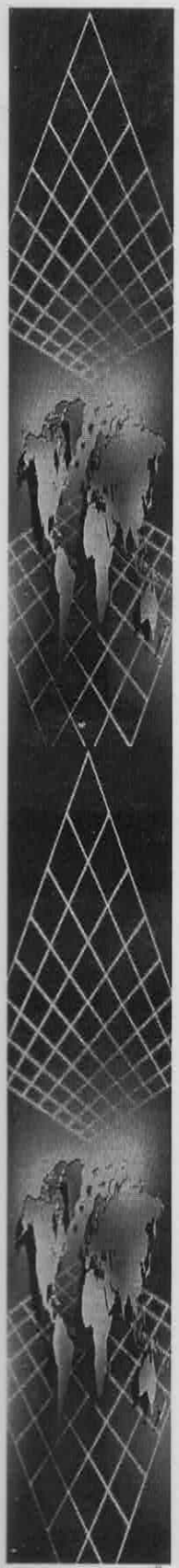
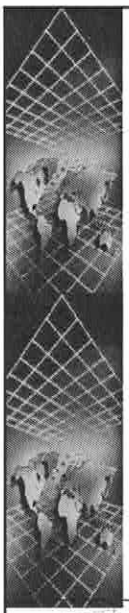


Global Knowledge Interchange Seminar

Michel **Biezunski (mb@infoloom.com)**,
Steven R. **Newcomb (srn@coolheads.com)**
Joel **Sachs (jsachs@cesdis40.gssc.nasa.gov)**

Paris, 23-24 January 2001
CNIT, Espace Elec : EDF

InfoLoom, Inc.
and
Coolheads Consulting
and
Joel Sachs

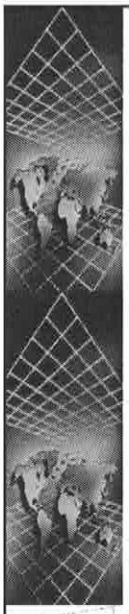


Global Knowledge Interchange Seminar

Michel **Biezunski**, (InfoLoom) mb@infoloom.com
Steven R. **Newcomb** (Coolheads Consulting), srn@coolheads.com
Joel **Sachs** (NASA), ottawajoel@aol.com

Paris, January 23-24, 2001
CNIT, Espace Elec: EDF

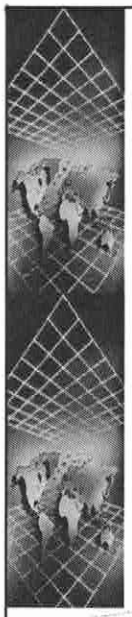
InfoLoom.Inc.
Coolheads Consulting and Joel Sachs



DAML, Knowledge Markup, and the Semantic Web

Joel Sachs
Goddard Earth Sciences and Technology Center
NASA Goddard Space Flight Center

InfoLoom.Inc.
Coolheads Consulting and Joel Sachs



Underlying Thesis of this Presentation

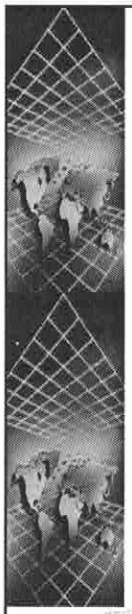
- The semantic web is evolving out of the efforts of three (not entirely distinct) communities.
  - The W3C: RDF and RDF Schema
  - The hypertext community: Topic Maps
  - The AI community: DAML
- To understand the semantic web, we should understand the motivations of these communities.
- To understand DAML, we will look at the subfields of AI that are informing DAML development; the goals of the DAML program; and some tentative DAML architectures.

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs

Overview of Presentation

- Background
  - Knowledge Representation
  - Agent oriented computing
- The concept of knowledge markup
- RDF, RDFS
- DAML
- Some DAML Examples and Architectures

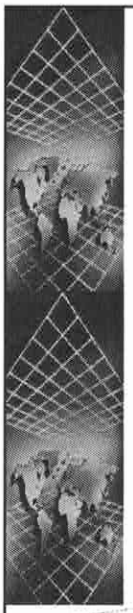
InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



A History of Knowledge Representation

- Knowledge representation (KR) is the branch of artificial intelligence (AI) that deals with the construction, description and use of ontologies.
  - How do we model a domain for input into the machine?
- *Ontology* is the branch of philosophy that answers the question "what is there?"
  - Some big names in ontology: Parmenides, Plato, Aristotle, Kant, Pierce, Husserl
- For a program to reason, it must have a conceptual understanding of the world. This understanding is provided by us. Thus we have to answer questions that we've been considering for several thousand years.
- Today, in computer science, an ontology is typically a hierarchical collection of classes, permissible relationships amongst those classes, and inference rules

InfoLoom, Inc.
Coolheads Consulting and Joel Sachs



State of the Art of AI/KR

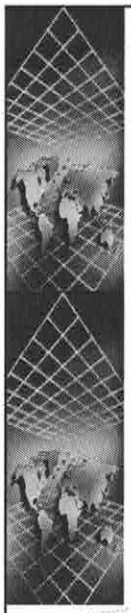
- 1970: Semantic Networks
  - Concepts are represented by nodes.
  - Relationships between concepts are represented by arcs.
- 1980: Frame Based Systems
  - Much more meta-data attached to objects.
    - Class, Sub-class hieracrchies
    - Attributes, facets, etc.
  - Took us a long way.
    - eg, first expert systems built with kee.
  - Still something missing: a need for rules, variables, quantifiers, etc. is recognized

InfoLoom, Inc.
Coolheads Consulting and Joel Sachs

State of the Art of AI/KR (contd.)

- 1990: Inference rules added
  - Modern Knowledge base systems.
  - A fairly rich subset of 1st order logic
  - recurring research issue: expressiveness vs. efficiency
    - It's easy to add a feature that takes you from one computational complexity level to another. (eg, adding negation kills monotonicity, which means complicated truth maintenance must be implemented.)
    - We will see this again.

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



Some Issues in Knowledge Engineering

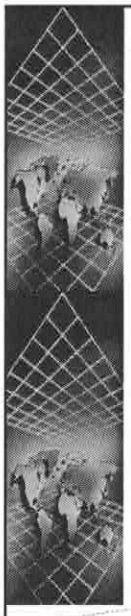
- How to represent, and reason about temporal objects.
- How to represent, and deal with change.
  - (This has been a problem since the pre-Socratics.)
- Need to be careful with inheritance.
  - Distinction between necessary and *non-necessary* properties.
    - eg, student *isa* person.
    - But a person has the property of always being a person.
    - Obviously, the class *student* does not inherit this property.

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs

The Agent Computing Paradigm

- The old way of thinking about computer programs: a program
  - » begins executing
  - » takes input
  - » gives output
  - » finishes executing
- The new way: programs
  - » interact with each other
  - » are always active
  - » should be robust (ie, able to deal with the unexpected)

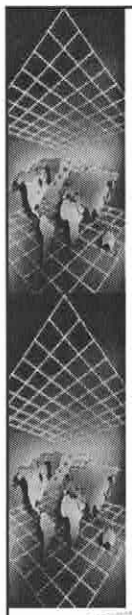
InfoLoom.inc.
Coolheads Consulting
and Joel Sachs



From Agents to Knowledge Markup

- Almost everything we need to know is on the web.
- What a great resource for agents!
- But ... Agents don't understand web pages.
  - – Natural Language processing is too hard for computers, and will remain so for a long time.
- The solution: Knowledge Markup.

InfoLoom.inc.
Coolheads Consulting
and Joel Sachs



Knowledge Markup in a Nutshell

- A web page describes objects.
  - Datasets, human beings, services, items for sale, etc.
- The semantics of an object are defined by the place it occupies in some domain ontology.
- The basic idea of knowledge markup is to use XML to markup a web page according to the location its objects occupy in the ontology.
- Essentially, knowledge markup is knowledge representation done in XML.

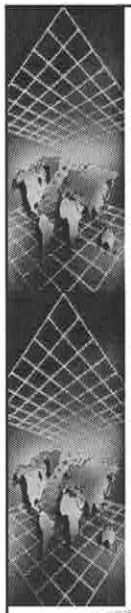
InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs

Generic Knowledge Markup Document

<ontology>
Some URLs
</ontology>

A collection of statements of the form:
<Class>
X
</Class>
<Relationship>
(X,Y)
</Relationship>

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



Benefits of KM

- Agents can parse a page, and immediately understand its semantics.
  - No need for natural language processing.
- Searches can be done on concepts. The inheritance mechanisms of the back-end knowledge base obviate the need for keywords.
- Data and knowledge sharing.

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs

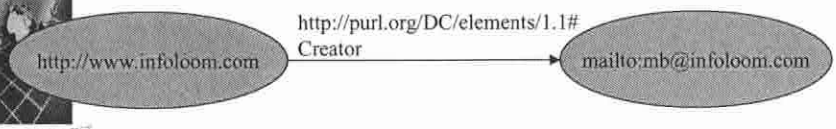
Knowledge Markup Example (Hypothetical)

- You ask the system "Show me all universities near the beach."
- The UCLA page doesn't say anything about the beach, but it does say (through knowledge markup) that it's near the Pacific Ocean.
- UCLA makes use of a geography ontology which includes the rule "Ocean(x) →hasBeaches(x)".
- When your search agent parses the UCLA page, it loads in the relevant ontologies, deduces that UCLA is near the beach, and returns the page.

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs

Resource Description Framework (RDF)

- RDF was conceived as a way to wrap metadata assertions (eg Dublin Core information) around a web resource.
- The central concept of the RDF data model is the *triple*, represented as a labeled edge between two nodes.
- The subject, the object, and the predicate are all resources, represented by URIs



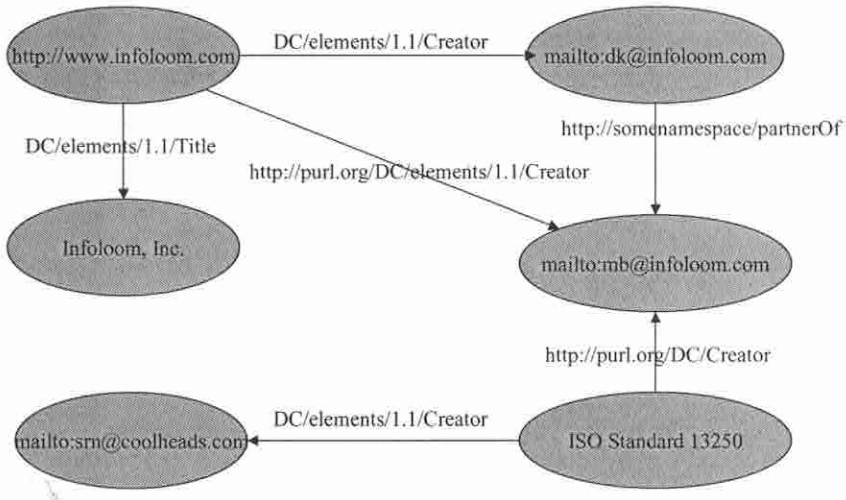
InfoLoom, Inc.
Coolheads Consulting
and Joel Sachs

RDF Data Model (Contd.)

- We say that <subject> has a property <predicate> valued by <object>.
- A resource may have more than one value for a given property.
- Objects may be valued by literals (instead of resources).
- Triples can be chained together, with the object of one triple being the subject of another.

InfoLoom, Inc.
Coolheads Consulting
and Joel Sachs

RDF Data Model (Contd.)



RDF Model: Reification

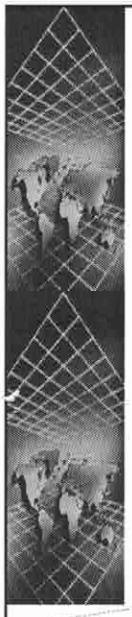
- Reify: *To regard or treat (an abstraction) as if it had concrete or material existence.* (Websters)
- Any RDF statement can be the object or value of a statement.
  - ie Graphs can be nested as well as chained.
- This allows us to make assertions about other people's statements.
  - eg "Joel Sachs believes that Michel Biezunski is the partner of Dianne Kennedy"

InfoLoom, Inc.
Coolheads Consulting
and Joel Sachs

The RDF Lexicon

- rdf:Resource
- rdf:Description
- rdf:Property
- rdf:Statement
  - rdf:subject
  - rdf:object
  - rdf:predicate
- rdf:Container
  - rdf:bag
  - rdf:sequence
  - rdf:alternative

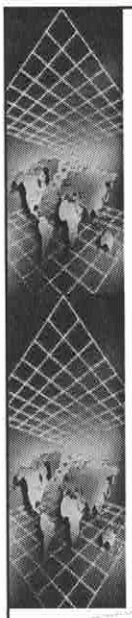
InfoLoom, Inc.
Coolheads Consulting
and Joel Sachs



RDF Syntax

- An XML syntax has been specified for RDF.
- An RDF document is a collection of assertions in **subject verb object** (SVO) form.
  - There are several accepted abbreviations.

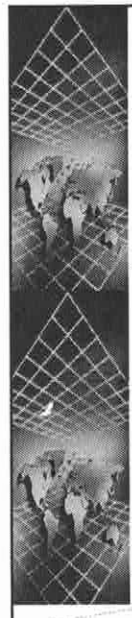
InfoLoom, Inc.
Coolheads Consulting
and Joel Sachs



RDF Syntax

```
<?xml version="1.0" encoding="UTF-8" ?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:dc="http://purl.org/DC/"
   xmlns:ns="http://someNameSpace/">
  <rdf:Description about="http://www.infoloom.com">
    <dc:Creator rdf:resource="mailto:mb@infoloom.com"/>
    <dc:Title> Infoloom, Inc. </dc:Title>
    <dc:Creator>
        <rdf:Description about="mailto:dk@infoloom.com">
          <ns:partnerOf  rdf:resource="mailto:mb@infoloom.com">
          </rdf:Description>
    </dc:Creator>
  </rdf:Description>
</rdf:RDF>
```

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs

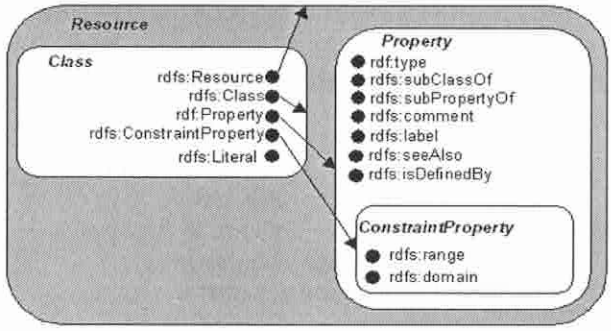


RDF Schema

- RDF Schema is a frame based language used for defining RDF vocabularies.
  - Introduces properties rdfs:subPropertyOf and rdfs:subClassOf
  - Defines semantics for inheritance and transitivity.
  - Introduces notions of rdfs:Domain and rdfs:Range
    - Also provides rdfs:ConstraintProperty
- We'll see an example when we look at DAML.

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs

RDF Schema Lexicon



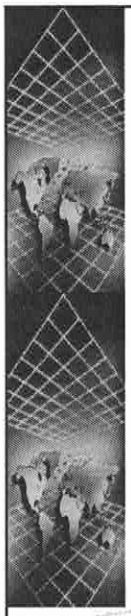
InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs

The Recapitulation of AI Research

- The last 2 years have seen a recapitulation of 40 years of AI history.
  - Data Structures →XML
  - Semantic Networks →RDF
  - Early Frame Based Systems →RDFS

As a mechanism for metadata encapsulation, RDFS works just fine. But it is unsuited for general purpose knowledge representation. This is where the AI community steps in, saying, essentially, "We know how to do this; please let us help."

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



The DARPA Agent Markup Language (DAML)

- A five year, $70 million research effort organised by the US Defense Army Research Project Agency, (the people who brought you the internet).
- Goal: To enable software agents to dynamically identify and understand information sources, and to provide semantic interoperability between agents.
- Activities:
  - Language Specification
  - Knowledge Annotation Tools
  - Construction of DAML aware multi-agent systems
    - The purpose of this last activity is to overcome a "chicken and egg" problem. The semantic web derives its utility from having many sites involved; but no one wants to get involved until a strong utility has been demonstrated.

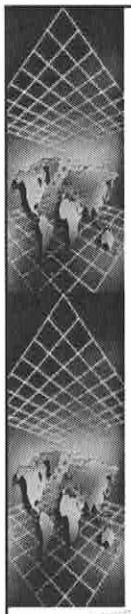
Advanced

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs

DAML (Contd.)

- The DAML research tasks have been distributed amongst 16 US research teams.
  - A veritable "who's who" of the active AI community.
    - Pat Hayes, Debbie McGuiness, Richard Fikes, Drew McDermott, Tim Finin, etc.
  - One of the teams is from the W3C (includes Tim Berners-Lee, Dan Conolly, Ralph Swick, etc.)

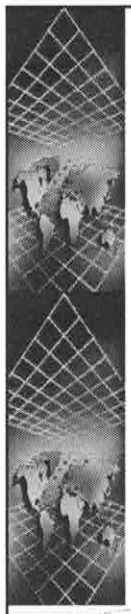
InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



DAML + OIL

- The most recent release of DAML incorporates the European Community's *Ontology Interface Layer* (OIL).
  - a product of the "Joint US/EU Committee on Agent Markup Languages."
- Written in RDFS
- This is a full fledged ontology modeling language.
  - Inference rules are due in the next release.

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs

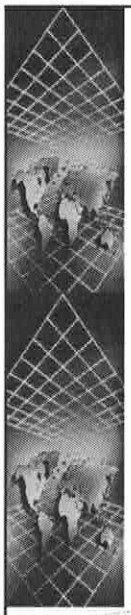


The best way to learn DAML is to read DAML

- Here we spend several minutes looking at an example DAML ontology, which will be passed around.
- Found at

http://www.daml.org/2000/12/daml+oil-walkthru.html

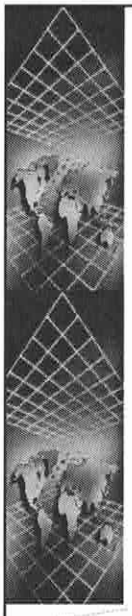
InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



Issues Facing DAML: Weak RDF semantics

- The semantics of RDF are ill-understood, and much controversed.
  - What does it mean for two classes to be equal?
  - What the heck is an rdf:container?
  - Part of the problem is a lack of clarity in the underlying web standards.
    - eg, the notion of URI identity.
- But the semantics of DAML must be clear. (Otherwise, no semantic interoperability.)
- Maybe the RDF-Topic Map Convergence will help clear things up.

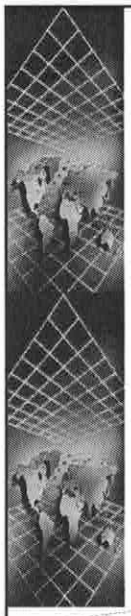
InfoLoom.Inc.
Coolheads Consulting and Joel Sachs



Issues Facing DAML: Legacy Data

- A need to datamine the legacy data, to determine appropriate DAML tags.
  - Work by Helka Folch, R. Holowczak
- Structured data is much easier to deal with than unstructured data: "T2XML"
- Much of our data is stored in databases. We publish it by dynamically generating HTML or XML pages. We could just as easily generate RDF or DAML pages.
  - That is, representing legacy data in DAML might not be as big a problem as it at first seems.

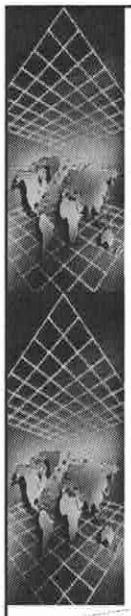
InfoLoom.Inc.
Coolheads Consulting and Joel Sachs



Issues Facing DAML: Need for Really Good Annotation Tools

- DAML is not meant to be read or written by human beings.
  - Humans will make assertions through intuitive user interfaces, which will generate the appropriate DAML markup.
- In fact, the markup should "fall out" of the activity of building a web page.
  - This requires some thought.

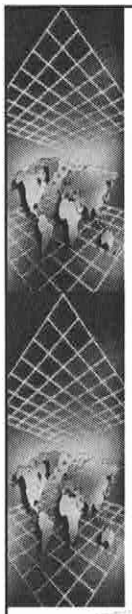
InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



DAML Example 1: Focused Crawling

- Special purpose search engines will increasingly replace all-purpose engines.
- The notion of an all-purpose search engine is yielding to that of special-purpose engines.
- Such engines do not want to index irrelevant pages.
- Current "focused crawling" techniques employ heuristics based on text mining, and collaborative filtering.
- A cleaner approach would be for web sites to describe themselves with DAML.
  - An entire site map could be expressed in DAML, along with metadata descriptions of each node in the map.
  - An agent would know precisely which of the site's pages are worth checking out.

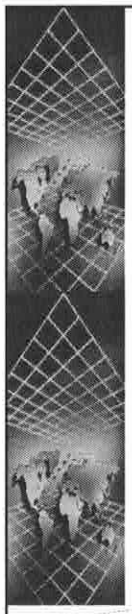
InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



DAML Example 2: Indexing the Hidden Web

- Search engines – google, infoseek, etc. – work by constantly crawling the web, and building huge indexes, with entries for every word encountered.
- But a lot of web information is not linked to directly. It is "hidden" behind forms.
  - eg www.allmovies.com allows you to search a vast database of movies and actors. But it does not *link* to those movies and actors. You are required to enter a search term.
- A web-spider, not knowing how to interact with such sites, cannot penetrate any deeper than the page with the form.

InfoLoom.Inc.
Coolheads Consulting and Joel Sachs

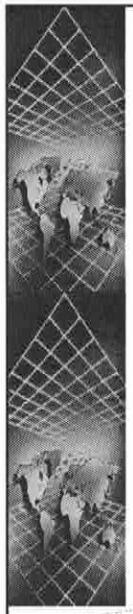


Indexing the Hidden Web (Contd.)

- Now imagine that allmovies.com had some DAML attached, which said

  "I am allmovies.com. I am an interface to a vast database of movie and actor information. If you input a movie title into the box, I will return a page with the following information about the movie: ... If you input an actor name, I will return a page with the following information about the actor: ..."

InfoLoom.Inc.
Coolheads Consulting and Joel Sachs



InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs

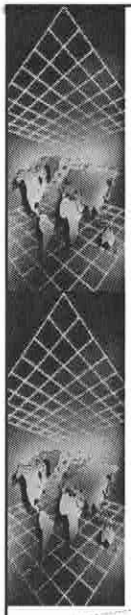
Indexing the Hidden Web (Contd.)

- A DAML aware spider can come to such a page and do one of two things:
  - If it is a spider for a specialized search engine, it may ignore the site altogether.
  - If not, it can say to itself: "I know some movie titles. I'll input them (being careful not to overwhelm the site), and index the results (and keep on spidering from the result pages).
- At the least, the search engine can record the fact that "www.allmovies.com/execperson?name=x" returns information about the actor with name x.

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs

DAML Example 3: An Environmental Legal Information System

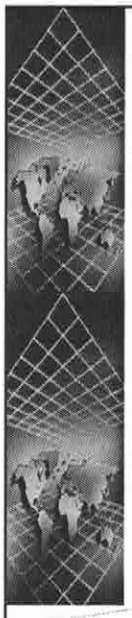
- The goal: interoperate remote sensing and environmental law databases.
- Sample query: Click on an environmental treaty, and ask "What remote sensing data do we have that can help in monitoring compliance of this treaty?"
- The problem: We can't expect the metadata attached to a particular remote sensing dataset to anticipate all queries to which it might be relevant. Reasoning must be done to determine which datasets to return.



DAML Example 4: Knowledge Sharing/Corporate Memory

- Our problem: NASA is huge, and IT practioners don't know what their colleagues are up to.
  - The wheel often gets reinvented.
- Our plan:
  - Build an ontology which captures the IT work being done at NASA.
  - Mark up projects, toolkits, algorithms, etc. according to this ontology.
  - Harvest the information with DAML aware web-crawlers.
  - Build DAML aware query agents.

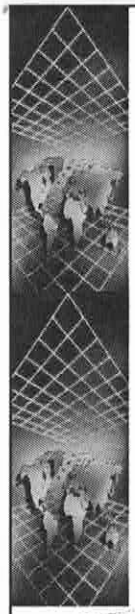
InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



DAML Example 4 (Contd.)

- Scientists should be able to tell the query agent the current form of their data (e.g. raw satellite images), their desired output (eg Time Series Forecasts), and get back the series of available tools necessary to perform the transformation.
- We also have a chicken and egg problem here.
  - Research teams don't want to invest time in yet another knowledge technology.
  - So we'll do it for them. We're selecting 20-30 divers projects at Goddard; we will interview the computer scientists, and mark up their efforts.

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



DAML Example 5: ittalks.org

- ittalks.org will be a repository of information about information technology (IT) talks given at universities and research organization across America.
- A user's information (research interests, schedule, constraints, etc.) will be stored on their personal DAML page.
- When a new talk is added, the personal agents of interested users will be notified.
- The personal agents will determine, based on schedule, driving time, more refined interest specifications, etc, whether or not to inform the user.

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs

ittalks.org (Contd.)

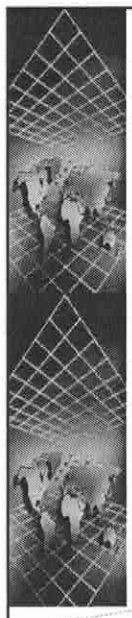
- Example Scenario
  - You are going to be in Boston for a few days. You enter this in your schedule, and you are automatically notified of several talks, at several Boston universities, that match your interests. You select one that you would like to attend. You get a call on your cell-phone letting you know when it is time to leave for the talk.

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs

The Road Ahead

- Enormous synergy between KM, ubiquitous computing, and agents.
  - Start Trek, here we come.
- The concept is clear, but many details need to be worked out.
- Semantic Web systems can be built incrementally.
  - Start small. Even a very modest effort can massively improve search results.

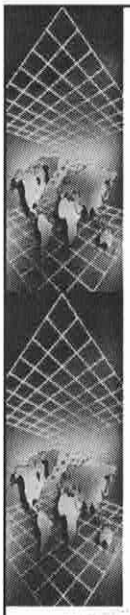
InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



Bibliographic Resources

- www.agents.umbc.edu
- www.semanticweb.org
- http://www.ladseb.pd.cnr.it/infor/Ontology/ontology.html !
- www.oasis-open.org/cover !
- www.daml.org
- mail: www-rdf-logic-request@w3.org
  - Subject: subscribe
- mail: majordomo@majordomo.ieee.org
  - In body: subscribe standard-upper-ontology
- RDF Tutorial, Pierre-Antoine Champin www.710.univ-lyon1.fr/~champin/rdf-tutorial

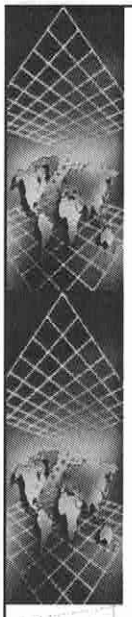
InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



Bibliographic Resources (contd.)

- Brickley, D. Guha, R.V. "Resource Description Framework Schema Specification 1.0" www.w3.org/TR/rdf-schema
- Decker, S. Melnik, S et al. "The Semantic Web: The Roles of XML and RDF" *IEEE Internet Computing,* September/October 2000
- Folch, H. Habert, B. "Constructing a Navigable Topic Map by Inductive Semantic Acquisition Methods" *Proceedings of Extreme Markup Languages 2000.*
- Freese, E. "Using Topic Maps: for the representation, management and discovery of knowledge" *Proceedings of XML Europe 2000.*
- Freese, E. "Topic Maps vs. RDF" *Proceedings of Extreme Markup Languages 2000.*
- Hefflin, J. Hendler, J. "Semantic Interoperability on the Web" *Proceedings of Extreme Markup Languages 2000.*
- Rath, H.H. "Topic Maps Self-Control" *Proceedings of Extreme Markup Languages 2000.*
- Stein, L. Connolly, D. McGuinness, D. "Annotated DAML Ontology Markup" www.daml.org/2000/10/daml-walkthru

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



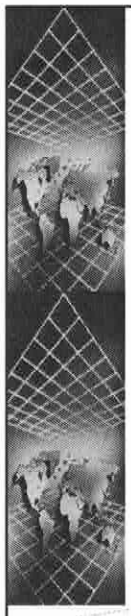
The Adolescence of XML

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs

A common fallacy: DTD is API

- The fallacy is: the structure of an XML resource should also be the API to the information it contains.
  - Trying to make the element structure also be the API makes it impossible to have both a good interchange structure and a good API. The attempt introduces inefficiency and invites unreliability of information interchange.
- The Document Object Model (DOM) is an API to the generic structure of XML resources. It is not and can never be the API to the information sets conveyed by all vocabularies.
  - If, e.g., the XLL vocabulary's functionality gets built into the DOM, what vocabulary's functionality *shouldn't* be built into the DOM?
    - No committee can possibly do all this work!

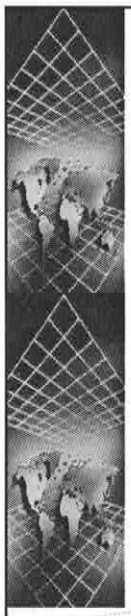
InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



Interchange syntax model is a contract

- DTD is a contract between
  - **information creators**
  - **information consumers**
  - applications developers
- DTD enhanced with type checking, lexical typing, etc., is a more detailed contract between the same players

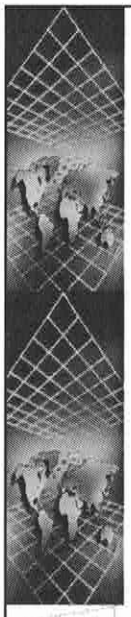
InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



Two sides of one coin

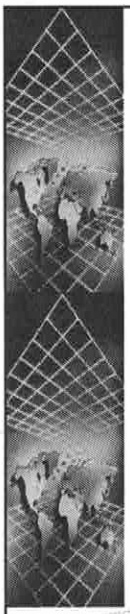
- The interchange syntax model and the abstract API are two aspects of the same information set:
  - Syntax model = consensus about the interchange format of the information set
  - Abstract API = consensus about the abstract properties of the information set

InfoLoom, Inc.
Coolheads Consulting
and Joel Sachs



Flexible Checking: the notion of Information Architecture

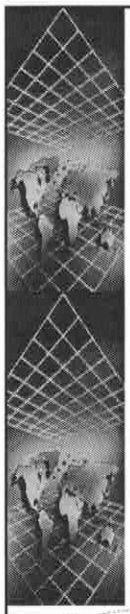
InfoLoom, Inc.
Coolheads Consulting
and Joel Sachs



Architecture Definition Document

- Documentation for an XML architecture:
  - meta-DTD (syntax rules)
  - Natural language documentation of the semantics of each form
- Can be used as a specification by programmers building architectural applications
- Can be used by document creators as a guide to proper usage

InfoLoom, Inc.
Coolheads Consulting
and Joel Sachs



What Makes an Architecture?

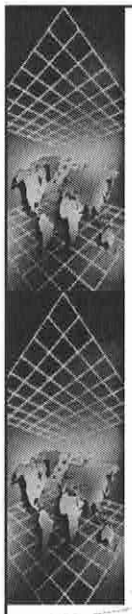
- Point of View: "base" and "client" roles
- "Client" inherits (is derived) from "base"
- Client DTDs inherit from base architectures
- Client meta-DTDs inherit from base architectures
  (architectures can be derived from other architectures)
- DTDs can be used as base architectures

InfoLoom, Inc.
Coolheads Consulting
and Joel Sachs

Architectures are about...

- Using constructs with predefined semantics as templates for specific element types
- Reusing software components that are built for defined architectures
- Inheriting semantics from more than one architecture at a time
- Accommodating diverse user groups while maintaining control at a general policy level

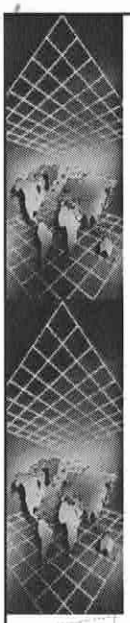
InfoLoom.inc.
Coolheads Consulting
and Joel Sachs



The Value of Architectures

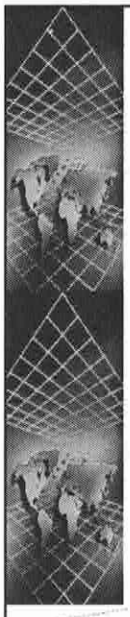
- Architectures allow user groups to
  - agree on common functionality while not abandoning group-specific semantics
  - use their own terminology
  - integrate architectural elements with group-specific elements (or use multiple architectures) separably, but in a single source document (customized data views)
  - use standardized architectural software

InfoLoom.inc.
Coolheads Consulting
and Joel Sachs



Challenges in Global Knowledge Interchange

InfoLoom, Inc.
Coolheads Consulting
and Joel Sachs



Be a smart shopper

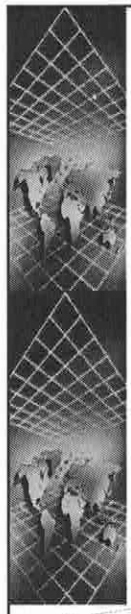
- Your organization accumulates knowledge every day.
- What is the lifecycle of knowledge?
- What causes the loss of knowledge?
- Purchasing decisions

InfoLoom, Inc.
Coolheads Consulting
and Joel Sachs

Is the Web designed for knowledge management?

- What does a Web address mean?
- Telecom etc. systems vs. information owners.
- Semantic web: A web for information owners?

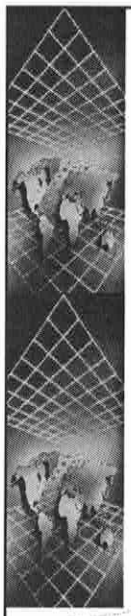
InfoLoom.Inc.
Coolheads Consulting and Joel Sachs



Names

- What is a name?
  - A property of a named thing?
  - A property of a network?
  - An indicator of semantics?
  - A resource?
  - An address?
  - A label?
  - An identifier?

InfoLoom.Inc.
Coolheads Consulting and Joel Sachs



Addresses

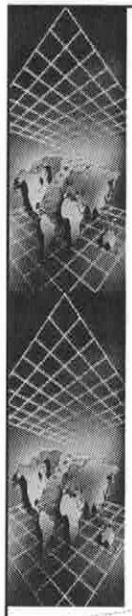
- What is an address?
  - A property of an addressable thing?
  - A property of a network?
  - An indicator of semantics?
  - A resource?
  - A name?
  - An identifier?

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs

Resources

- What is a resource?
  - An addressable thing?
  - The content of an addressable thing?
  - The meaning of an addressable thing?

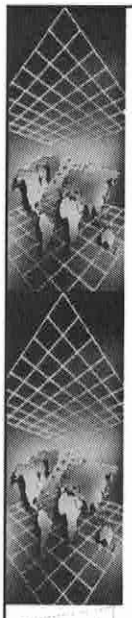
InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



Concepts: What It's All About

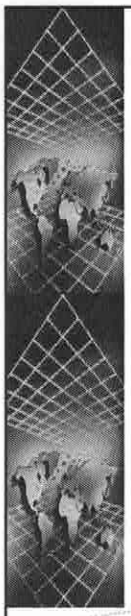
- Need a foundation concept of concept.
- Need a discipline for associating concepts with resources.
- Need to give specific roles to specific kinds of resources.
  - Name resources
  - Address resources
  - Concept-defining resources

InfoLoom, Inc.
Coolheads Consulting
and Joel Sachs



XML Vocabularies

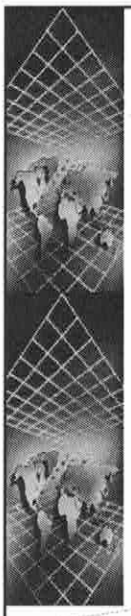
InfoLoom, Inc.
Coolheads Consulting
and Joel Sachs



XML "Namespaces"

- XML "Namespaces" are vocabularies. The XML "Namespace" recommendation is a step on the road toward interoperability for XML messages.
- A "namespace" amounts to an abstract "place" where there is a list of element type names (tag names) and/or attribute names.
- URIs specify the namespaces in use.
- There is no requirement that the specified URI is valid, much less that the indicated resource conforms to any sort of specification.
- XML "Namespaces" provide a way for names to be guaranteed to be unique, and *that's all.*

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



XML "Namespaces" and expectations

- In some sense, an XML resource that uses the names of an XML "namespace" must inherit from it expectations as to the meaning and conventional use of each of the names.
  - (Right? Otherwise, why use it at all?)

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs

Do namespaces help with interchange?

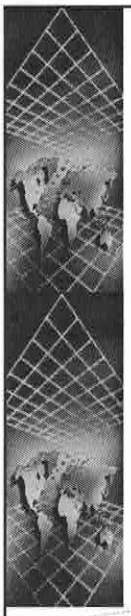
```
<hp:hperm
 xmlns:hp='http://www.gov.ng/hp.x'>
<hp:thone>UDLINGBLON</hp:thone>
<hp:kallow>29</hp:kallow>
<hp:spec>GUINEA FOWL</hp:spec>
<hp:date>2000 3 9</hp:date>
</hp:hperm>

(Vocabulary in the HP namespace:
 hperm, thone, kallow, spec, date)
```

InfoLoom, Inc.
Coolheads Consulting and Joel Sachs

XML "Namespaces:" Unresolved issues

- How to express a namespace so it can be shared? What is the list of names?
- How to write processing software for a namespace?
- How to determine whether software for a namespace works according to the expectations of other users of the namespace?
- How to determine whether an XML resource conforms to the syntactic and semantic requirements of the namespace?
- How to determine, when information interchange fails, whose software is at fault? (The software that created the XML resource, or the recipient application?)
  - Accountability is vital to interchange.

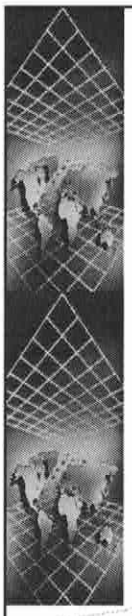
InfoLoom, Inc.
Coolheads Consulting and Joel Sachs



XML Vocabularies in open environments

- Ideally, an XML resource is self-describing. Since many XML resources use the same vocabularies, it's efficient to describe them in terms of the vocabularies they use.
- Anybody who receives a well-described XML resource should be able to interpret it accurately.
- Anybody should be able to create an XML resource that uses a vocabulary correctly, so that its recipient will interpret it accurately.
- Vocabularies should be able to support entire industries and areas of human endeavor, in open, multivendor environments.
- Vocabularies should offer huge advantages in efficiency and reliability.

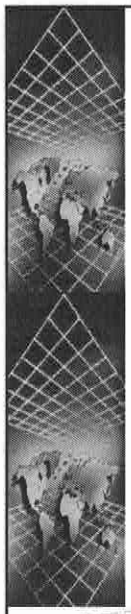
InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



XML Vocabularies in closed environments

- Closed syndicates and would-be cartels need to resolve the same issues, so that their XML messages will interoperate.
- It's extremely inefficient for each syndicate to invent the methodologies and tools for guaranteeing reliable vocabulary-based interoperability.
  - It's also a net contraction in the noosphere of the syndicate. Where to find technical expertise? How to maintain it? Etc.
- Enlightened self-interest demands that the same methodologies and tools that support open interoperability be used internally.
- Vocabularies should offer huge advantages in efficiency and reliability.

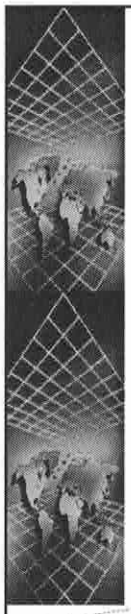
InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



Methodologies and Tools for Vocabularies

- Vocabularies can be used to make XML resources fully self-describing, fully interchangeable and fully interoperable, down to the last syntactic and semantic feature.
- This can be accomplished using existing W3C and ISO recommendations and standards, all from the XML and SGML families of recommendations and standards.
  - Alternatively, the same principles could be applied using different modeling syntaxes, purpose-built for the Web.
  - ...but if it can be done without reinventing everything, why bother?

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



How to document vocabularies?

- It would be great to be able to document vocabularies more effectively than we can now.

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs

Which constructs are the comments about?

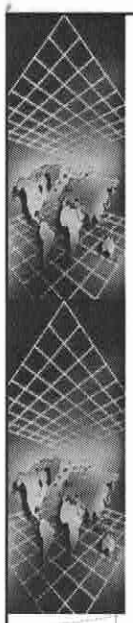
```
<!-- Indigenous person's hunting permit -->
<!ELEMENT hperm  (thone, kallow, spec, date)>
<!ELEMENT thone  (#PCDATA)> <!-- person's name -->
<!ELEMENT kallow (#PCDATA)> <!-- kill allowance -->
<!ELEMENT spec   (#PCDATA)> <!-- species -->
<!ELEMENT date   (#PCDATA)> <!-- expiration date -->
  <!-- lexical model: YYYY MM DD -->

<hperm>
<thone>UDLINGBLON</thone>
<kallow>29</kallow>
<spec>GUINEA FOWL</spec>
<date>2000 3 9</date>
</hperm>
```

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs

Documenting vocabularies

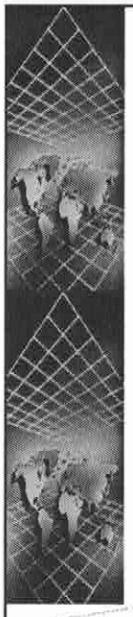
- Topic maps are an extremely powerful way of documenting DTDs.
  - ...but that's coming later in this talk.

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



Universal Addressing

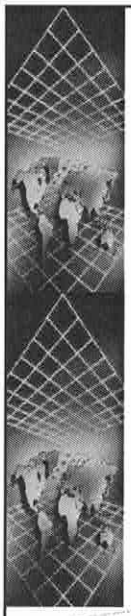
InfoLoom.inc.
Coolheads Consulting
and Joel Sachs



Groves, Graphs and Queries

- The concept of Grove
- Topic Map Graph
- RDF Graph
- Specialized Query Languages

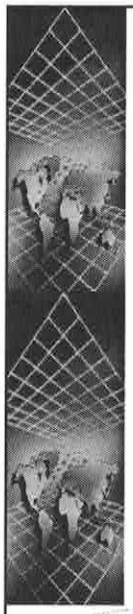
InfoLoom.inc.
Coolheads Consulting
and Joel Sachs



It's all about addressing

- External metadata (such as topic maps) must be associated with data by means of addresses.
- How can the value of external metadata be protected against losses due to technology change? In other words:
- How can addressing expressions be independent of technology?

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



Semantic modeling: emergent properties

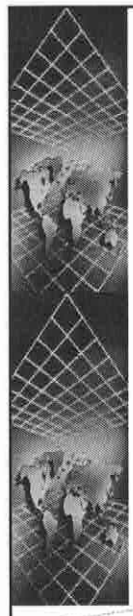
- Example of an "emergent" property: The property of being a target of an xlink (considering XLL as a vocabulary).
- All emergent properties of a vocabulary must be described clearly, comprehensively, unambiguously, and formally, because
  - accuracy and reliability are important.
  - the information is expected to be useful in multi-vendor application environments (if not, why inherit a vocabulary at all?).
  - implementation of vocabulary-specific applications must be done at reasonable cost.

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs

Semantic validation becomes a side-effect

- Computing an emergent property value often isn't possible without validating the interchanged information on which the computation is based.
  - For example, if an element that inherits from a vocabulary specifies a "start-time" attribute and an "end-time" attribute, we may intend that the duration of time between the start-time and the end-time be calculable and that it fall within a certain range (or at least be non-negative). In any case, we can't calculate the value of the "duration" property unless the start-time and end-time values exist and are amenable to calculation.

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs

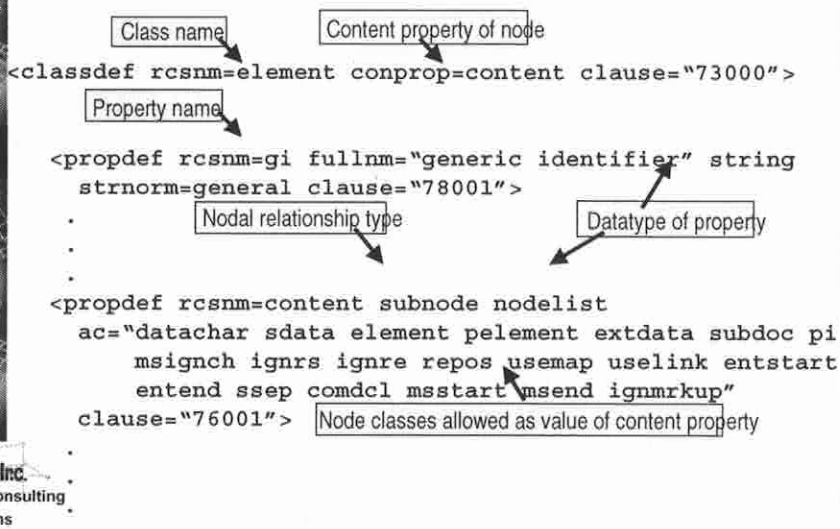


ISO standard property language

- It's called "Property Sets"
- A property set is an XML document that conforms to the ISO standard DTD for property sets.
- Already in commercial use; the software already works with XML.
- Every class of information component ("node"), and every property of every class, has a unique name.
- These names can be used in queries.
- This whole idea is often called "the Grove Paradigm." It's the basis of SGML processing, and the SGML Property Set aided the development of the DOM.

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs

Node Class Definition

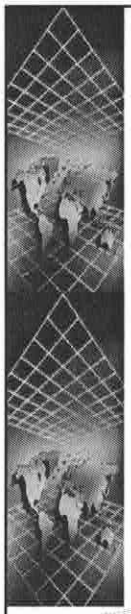


InfoLoom, Inc.
Coolheads Consulting and Joel Sachs

Designing XML Vocabularies

- Questions to ask:
  - Must certain semantic processing and validation operations be performed by all applications of this vocabulary?
  - Will more than one application have to deal with this vocabulary?
- If so, its syntactic requirements deserve to be made explicit in a DTD (or something like a DTD), and
- A property set (or other explicit Abstract API) defined for it will pay big dividends
  - in software reuse
  - in achieving widespread consensus about what the vocabulary really means
  - in determining what went wrong when vocabulary-mediated information interchange fails

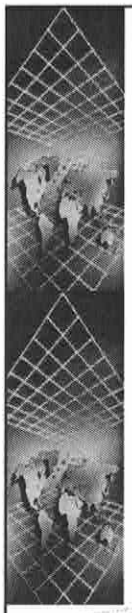
InfoLoom, Inc.
Coolheads Consulting and Joel Sachs



Groves go beyond XML

- Not all information can be efficiently represented as XML
- Conversion to XML can be expensive
- XML is continuously redefined and evolving
- XML is a property set
- APIs to meanings of schemas, DTDs, and vocabularies are Property Sets

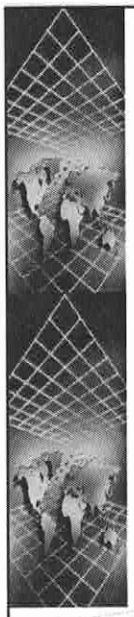
InfoLoom, Inc.
Coolheads Consulting
and Joel Sachs



Groves go beyond the DOM

- The world is more than documents
- DOM is evolving
- DOM is a Property Set
- Properties can be extended or retracted to suit any processing needs, current *or* future!

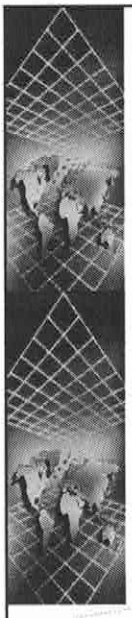
InfoLoom, Inc.
Coolheads Consulting
and Joel Sachs



Groves are future proof

- Grove paradigm intercepts enhancements to WWW
- Grove paradigm intercepts future extensions to W3C Recommendations
- Grove paradigm allows smooth evolution of software infrastructure
- Grove paradigm allows less-frequent abandonment of costly software

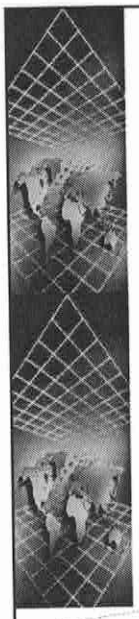
InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



Some definitions of "grove":

- A set of interconnected nodes conforming to ISO/IEC 10744:1997 Annex A.4.
- **Ready-to-run objects.** The application-internal information that is being created, edited, or used by an application.
- The form of information that is **the result of parsing an interchange document** (written in XML, for example).
- Information that is present in an interchange document, but which, for application programming, is structured differently in the grove. **The abstract API to the information set.**

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



But that's the DOM, right?

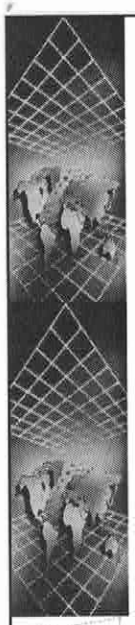
- No. The DOM is an API. A grove is a set of nodes -- no methods.
  - But it's true that an API to a grove of an XML document could serve the same purposes as the DOM.
  - And the DOM is an example of an interface to an XML grove.
- Groves are more powerful, general and sophisticated.

InfoLoom.inc.
Coolheads Consulting
and Joel Sachs

Use by reference

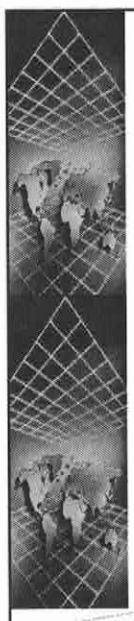
- Groves allow any component(s) of any resource(s) expressed in any notation(s) (XML or non-XML) to be transcluded into any components of any resources expressed in any notations (XML or non-XML),
- ...without changing any of the resources involved.
- Groves allow there to be a single authoritative copy of each datum, regardless of its notation.
- Groves allow each datum to be expressed in the most cost-effective notation.

InfoLoom.inc.
Coolheads Consulting
and Joel Sachs



Metadata

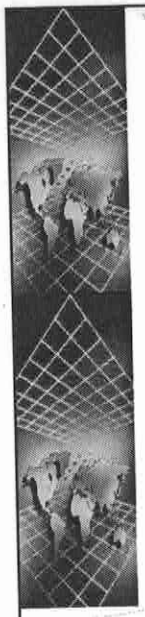
InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



Discovery is a hard problem

- Information is multidimensional, multimedia, multicultural, multilingual, context sensitive and multicontextual.
- The information you're looking for may require that you read and understand it yourself before you can recognize that it is what you need.
- There is an awful lot of information in your own place of business, let alone the Web. The overwhelming majority of it is not what you're looking for.
- Nobody knows what you're looking for but you. (And you yourself may not know.)

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



Metadata for discovery

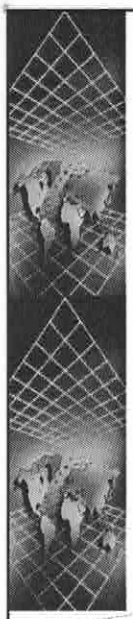
- Topic maps
- RDF
- Dublin Core, MARC, etc.

InfoLoom.inc.
Coolheads Consulting
and Joel Sachs

Undiscovery is a hard problem

- How to disqualify irrelevant information?
  - Saying what something is relevant to says nothing about what it's irrelevant to. Neither the absence nor the presence of, e.g., a keyword indicates irrelevance.
  - Need to define a whole universe of possible relevance in order to say what's irrelevant.
- How to disqualify obnoxious information?
  - Pornography vs. art.
  - Medical information (behaviors that risk AIDS, what a healthy breast is supposed to look like) vs. religion/culture
  - Self description is always misleading, usually deleterious to free speech.
  - Value judgments are protected speech, too.

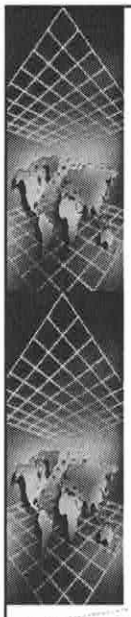
InfoLoom.inc.
Coolheads Consulting
and Joel Sachs



Metadata for undiscovery

- Topic maps
- RDF
- PICS
- Digital signature
- P3P

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



Metadata architectures

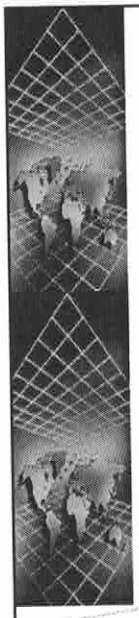
- Dublin Core
- ICE, CDF
- RDF
- PICS
- WebDAV

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs

Dublin Core

- MARC on a strict diet. (OCLC is in Dublin, Ohio.)
- Seeks "simplicity, semantic interoperability, international consensus, extensibility, modularity"

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



Dublin Core Metadata Element Set V. 1.1

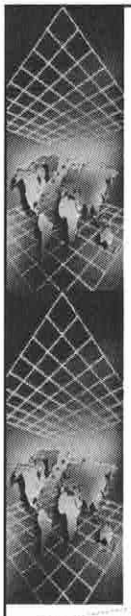
- Core elements:
  - title
  - subject (keywords)
  - creator
  - description (abstract, etc. of content)
  - publisher
  - contributor
  - date
  - type (genres, categories, etc.)
  - format (physical/digital manifestation)
  - identifier (unique string/number in some formal system)
  - source (derived from)
  - language (some natural language used in content)
  - relation (reference to related resource)
  - coverage (scope of the content)
  - rights (who owns what rights, or how to find out)

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs

ICE - Information & Content Exchange

- Designed to enable:
  - syndicated publishing networks
  - Web superstores
  - online reseller channels
- XML-based message protocol
  - exchange, update information assets
  - trust relationships (builds on P3P)
  - resellers/aggregators add value; consumers find what they want with less difficulty
- Names to drop:
  - Adobe, CNET, GCA, Microsoft, National Semiconductor, Sun Microsystems, Tribune Media Services, Vignette

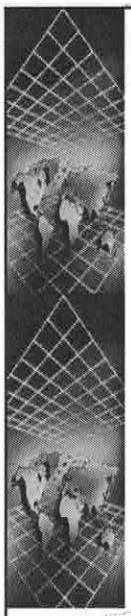
InfoLoom.Inc.
Coolheads Consulting and Joel Sachs



ICE, continued

- The ICE protocol defines the roles and responsibilities of syndicators and subscribers,
- defines the format and method of content exchange, and
- provides support for management and control of syndication relationships.
- ICE defines a vocabulary and a messaging protocol.
- Industry-specific vocabularies are expected to be mixed with it.
- "Payload" is an oft-heard word in ICE-land.

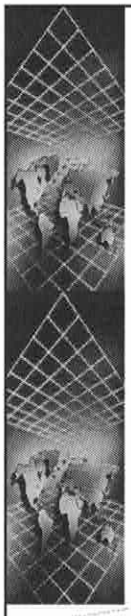
InfoLoom.Inc.
Coolheads Consulting and Joel Sachs



CDF, Channel Definition Format

- Channel Definition Format (CDF) specifies the operation of push channels.
- Like ICE, it defines a mechanism for scheduling delivery of encapsulated content.
- Unlike ICE, CDF has no notion of explicit subscription relationship management, asset management, reliable sequenced package delivery asset repair operations, constraints, etc.

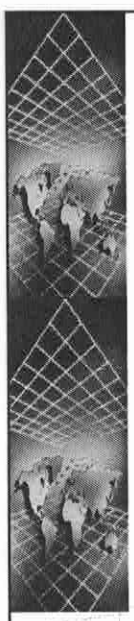
InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



PICS, Platform for Internet Content Selection

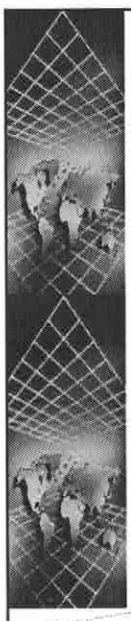
- A way of associating "labels" with internet resources.
- For use by programs running in conjunction with browsers that block access to internet resources.
- Originally for the protection of children's sensibilities, while allowing them otherwise-unimpeded access to the Internet (allowing the Internet to be used with minimal supervision, as a kind of TV-babysitter).
- Numeric values for categories.

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs

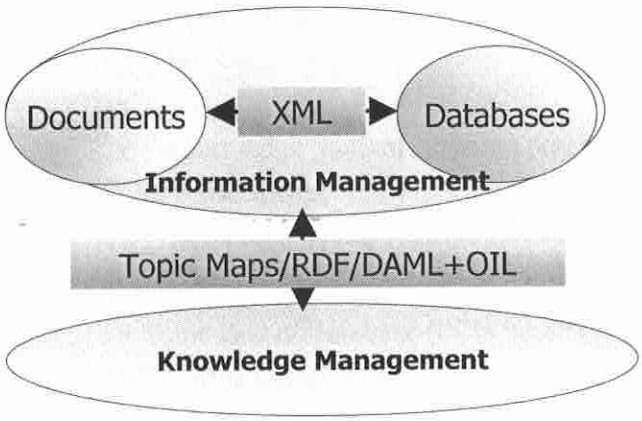


Finding Information

InfoLoom.inc.
Coolheads Consulting
and Joel Sachs



A New Bridge



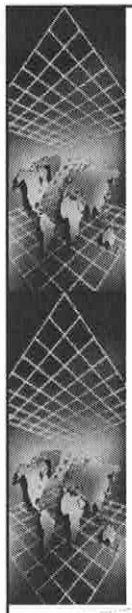
InfoLoom.inc.
Coolheads Consulting
and Joel Sachs

Enabling Global Knowledge Connectivity

- Strings (full text indexes, keyword indexes) are not sufficient for searching information.
- Structures (document types, schemas and taxonomies) are sometimes too constraining to qualify (categorize, describe) information.
- Information should be findable regardless of its format or platform of origin.

To find information, we need information *finding aids* that go beyond strings, work with all structures, and are platform-independent.

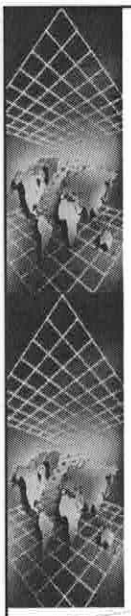
InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



How to make findable the information found in:

- Unstructured Sources
  - Create
    - Full text index
    - Subject index (natural language processing)
  - Query the index
- Structured Sources
  - Create
    - Some kind of Database (DOM for XML)
  - Query the Database

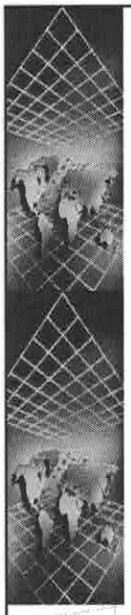
InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



Standardized Solutions

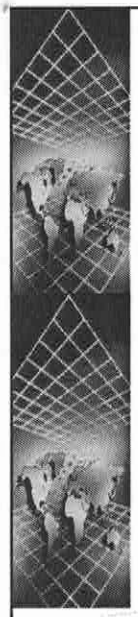
HTML	Web
XML	Text, Databases
RDF	Connectivity Layer
DAML+OIL	Knowledge Representation
Topic Maps	Finding Aids

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



Topic Maps

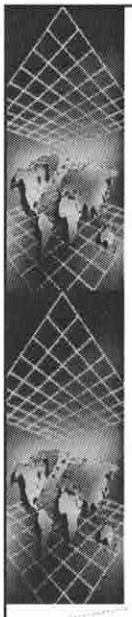
InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



Standard(s)

- ISO/IEC 13250:2000 Topic Maps (January 2000)
- Topicmaps.org XTM (first release: December 2000)
- Origin:
  - Davenport (1991-1993)
  - CApH (1993-1995)
  - ISO (1996-1999)

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs

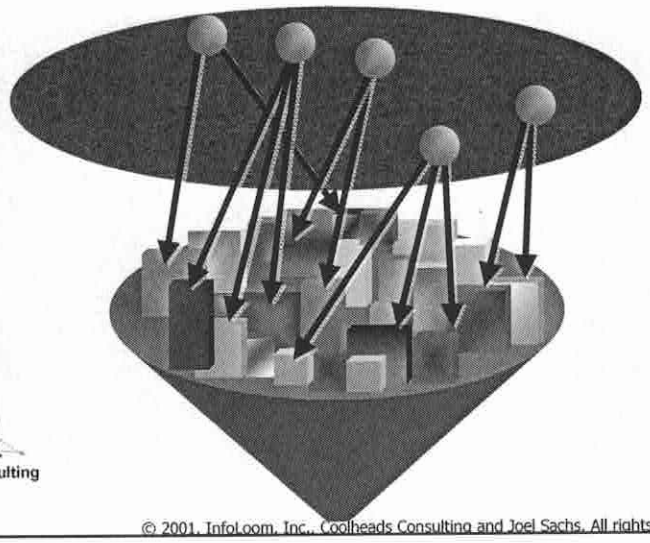


Topic maps at a glance

- Subjects, subject descriptors
- Topics
- Names
- Occurrences
- Associations
- Scopes

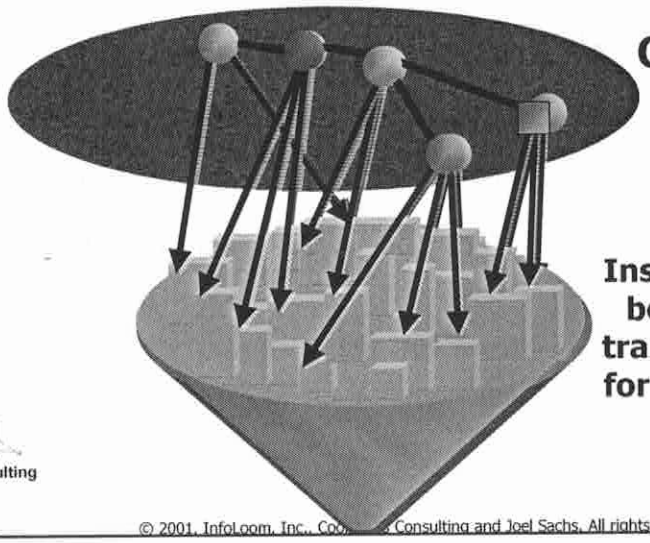
InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs

Topics as Information Hubs



InfoLoom.inc.
Coolheads Consulting
and Joel Sachs

User-defined semantics

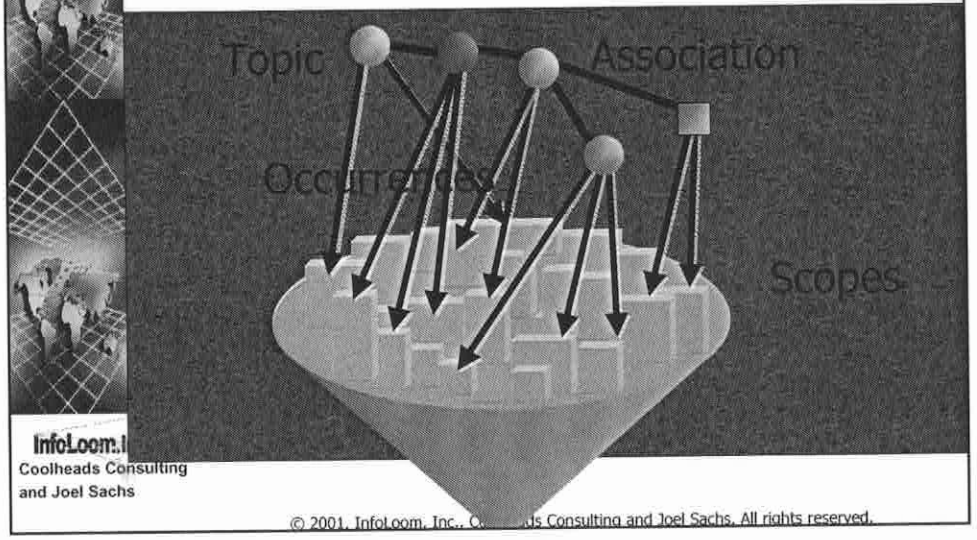


Outside layer

Inside view becomes transparent for the user

InfoLoom.inc.
Coolheads Consulting
and Joel Sachs

Topic map constructs

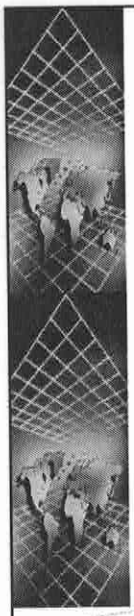


InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs

Topics

- Subject is essence of topic.
- Topics have Names.
- Topics have Occurrences.
- Can be associated with other topics.
  - Topics can be instances of other topics.

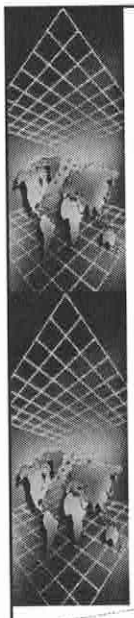
InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



Names

- A topic can have 0, 1 or several names.
- Base name and variant names (e.g., Display, Sort)
- Names have scopes
  - Scopes differentiate names by e.g., language, domain, usage, security, etc.
- A scope is a set of topics.

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs

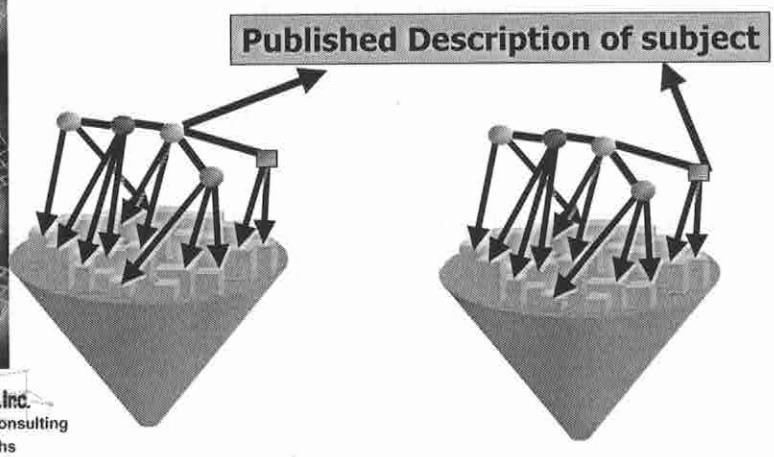


Occurrences

- Collections of resources that are relevant to a topic.
- Occurrences are grouped by class (e.g., depiction, etc.)
- Occurrences have scope.
- An occurrence can be a subject descriptor.

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs

Shared subject descriptors

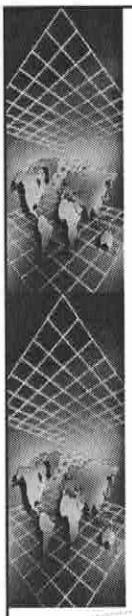


InfoLoom.inc.
Coolheads Consulting
and Joel Sachs

Subjects

- The subject is the abstract notion which is the invisible heart of every topic.
- The object that describes a subject is called a subject descriptor.
- Subject descriptors can be published, i.e., available for public consumption.

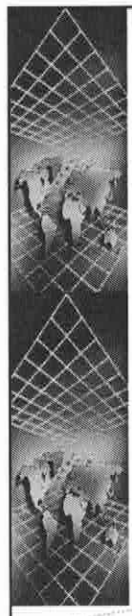
InfoLoom.inc.
Coolheads Consulting
and Joel Sachs



Subject descriptors: two senses

- A subject descriptor can be used 2 ways:
  - for the subject it describes, and
  - for the subject that is itself.
- You have to say which sense. Default is the subject it describes.

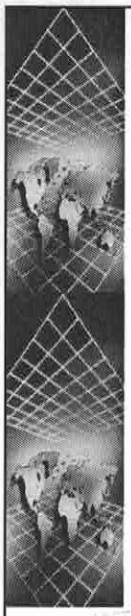
InfoLoom.Inc.
Coolheads Consulting and Joel Sachs



Merging topic maps

- Two topics that have the same subject descriptor in the same sense become a single resulting topic.
- A topic map can reference the topic maps with which it is supposed to be merged.

InfoLoom.Inc.
Coolheads Consulting and Joel Sachs

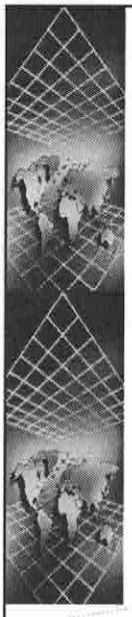


Topic Maps vs. RDFS

- Built from the top down.
  - First we define the topics, with pointers to resources.
  - Then we can make those pointers bidirectional to enable cross-referencing.
- Motivated by problems of electronic publishing.
- Everything is a topic.
- Intuitive

- Built from the bottom up.
  - We annotate a document with RDF.
  - Then we parse all the RDF pages and build an index.
- Motivated by problems of library science and knowledge sharing.
- Everything is a resource.
- Hard to understand

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs



Topic Maps vs. RDF (contd.)

- Using RDF for KR: DAML
- Using Topic Maps for KR: ?
  - Work by Eric Freese, Hans Holger Rath.

- The co-authors of RDF and the co-authors of Topic Maps are currently negotiating the merging of the two standards.

InfoLoom.Inc.
Coolheads Consulting
and Joel Sachs