# Enriching Documents in an Information Portal using Superimposed Schematics[*]

Shawn Bowers Lois Delcambre David Maier
OGI School of Science and Engineering
OHSU, Beaverton OR 97006, USA
{shawn, lmd, maier}@cse.ogi.edu

April 25, 2002

**Abstract**

Portals offer relevant information often with improved and customizable search. However, most portals concentrate on locating documents as opposed to enhancing their use. We present superimposed schematics which serve to enrich documents by adding a structured, conceptual guide for their contents. A schematic provides entity-relationship (E-R) style structures integrated with marks, where each mark holds an address to an excerpt in an underlying document. Schematics enable enhanced addressing of documents, conceptual navigation, and query all at "no additional cost," i.e., without modifying the base documents. We report on superimposed schematics and discuss their application to documents, such as the Appeal Decision of the USDA Forest Service, within the Adaptive Management Portal.

## 1 Introduction

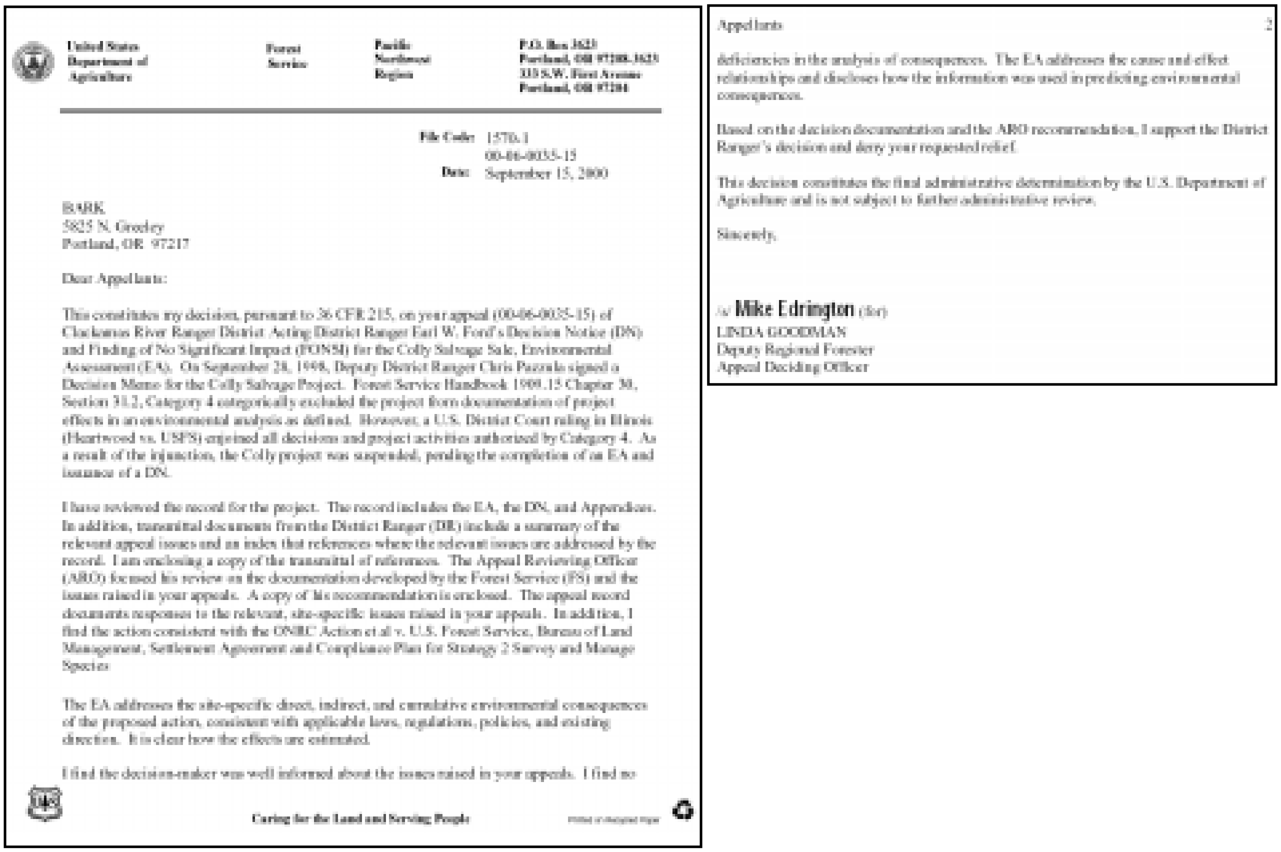
The actions of the USDA Forest Service, such as selling timber or issuing or denying special use permits, are officially documented in *Decision Notices*, *Records of Decision*, and so forth. Members of the public then have the right to file an appeal requesting that the decision be changed. When the appeal deadline has passed, a *reviewing officer* from the Forest Service normally considers the entire set of appeals for a given decision and makes a recommendation. Finally, the *deciding officer* makes a determination for each issue raised (in one or more appeals). Each *appeal decision*[1] is typically represented in two standard letters, one from the deciding officer and one from the reviewing officer (see Figure 1 for an example decision letter).

Although not easy to discern from looking at the documents, an appeal decision is comprised of a fairly standard set of information items, such as: (1) which decision is being challenged, (2) which appellants have filed (and on behalf of which organization(s)), (3) what issues were raised (across the set of appeals), and (4) what final determination was made for each issue. We observe that the standard items and relationships among them can be usefully represented in a *superimposed schematic* (Bowers et al., 2002), an entity-relationship (E-R) style schema (Chen, 1976) for superimposed information, as shown in Figure 2.

The key feature of superimposed schematics is that they can be populated with *marks*, which are integral to superimposed information in general (Delcambre et al., 1997; Delcambre and Maier, 1999; Delcambre et al., 2001a) in general), in addition to regular attribute values. Each mark represents and holds an address for an information excerpt in an underlying document (e.g., in a decision letter), as shown in Figure 3. For example, we see that a particular appeal is mentioned in the first paragraph of the decision letter and that

[*] This work supported, in part, by the National Science Foundation through grants EIA 9983518 and IIS 9817492. Information and examples concerning the Forest Service and appeal decisions were provided by Timothy Tolle of the USDA Forest Service.

[1] The Pacific Northwest Region of the Forest Service has appeal decisions online for twenty national forests dating back to 1997 at http://www.fs.fed.us/r6/plan/appeal.htm.



United States
Department of
Agriculture

Forest
Service

Pacific
Northwest
Region

P.O. Box 3623
Portland, OR 97208-3623
333 S.W. First Avenue
Portland, OR 97204

File Code: 1570-1
00-06-0035-15
Date: September 15, 2000

BARK
5825 N. Greeley
Portland, OR 97217

Dear Appellants:

This constitutes my decision, pursuant to 36 CFR 215, on your appeal (00-06-0035-15) of Clackamas River Ranger District Acting District Ranger Earl W. Ford's Decision Notice (DN) and Finding of No Significant Impact (FONSI) for the Colly Salvage Sale, Environmental Assessment (EA). On September 28, 1998, Deputy District Ranger Chris Pazzola signed a Decision Memo for the Colly Salvage Project. Forest Service Handbook 1909.15 Chapter 30, Section 31.2, Category 4 categorically excluded the project from documentation of project effects in an environmental analysis as defined. However, a U.S. District Court ruling in Illinois (Heartwood vs. USFS) enjoined all decisions and project activities authorized by Category 4. As a result of the injunction, the Colly project was suspended, pending the completion of an EA and issuance of a DN.

I have reviewed the record for the project. The record includes the EA, the DN, and Appendices. In addition, transmittal documents from the District Ranger (DR) include a summary of the relevant appeal issues and an index that references where the relevant issues are addressed by the record. I am enclosing a copy of the transmittal of references. The Appeal Reviewing Officer (ARO) focused his review on the documentation developed by the Forest Service (FS) and the issues raised in your appeals. A copy of his recommendation is enclosed. The appeal record documents responses to the relevant, site-specific issues raised in your appeals. In addition, I find the action consistent with the ONRC Action et al v. U.S. Forest Service, Bureau of Land Management, Settlement Agreement and Compliance Plan for Strategy 2 Survey and Manage Species

The EA addresses the site-specific direct, indirect, and cumulative environmental consequences of the proposed action, consistent with applicable laws, regulations, policies, and existing direction. It is clear how the effects are estimated.

I find the decision-maker was well informed about the issues raised in your appeals. I find no

Caring for the Land and Serving People

Printed on Recycled Paper

Appellants 2

deficiencies in the analysis of consequences. The EA addresses the cause and effect relationships and discloses how the information was used in predicting environmental consequences.

Based on the decision documentation and the ARO recommendation, I support the District Ranger's decision and deny your requested relief.

This decision constitutes the final administrative determination by the U.S. Department of Agriculture and is not subject to further administrative review.

Sincerely,

/s/ Mike Edrington (for)
LINDA GOODMAN
Deputy Regional Forester
Appeal Deciding Officer

Figure 1: The decision letter for an example appeal decision.

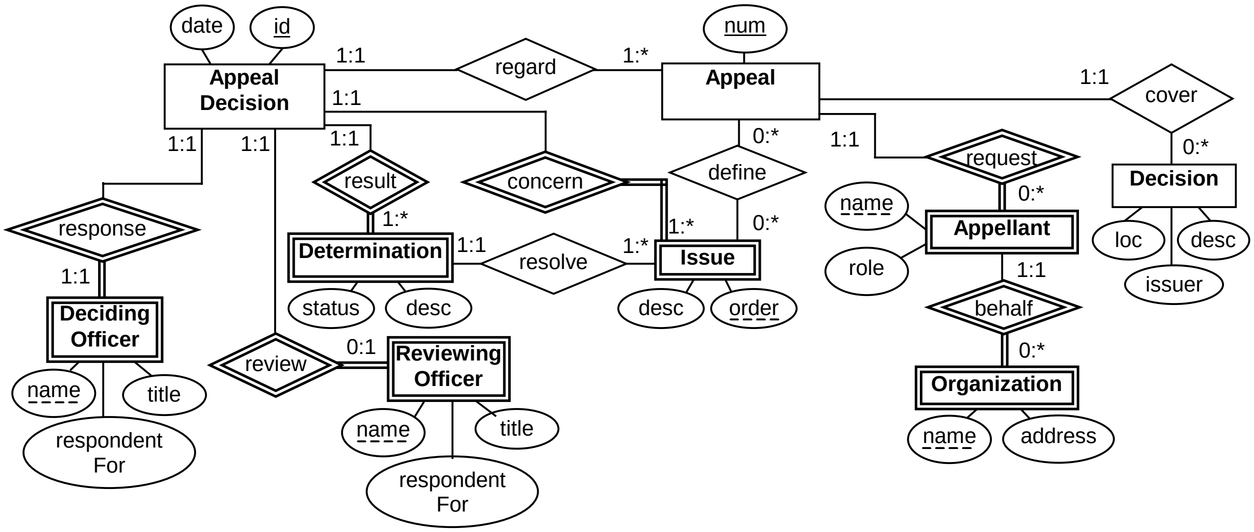


Figure 2: Conceptual structure of the standard information elements in an appeal decision.

the issue raised in that appeal is described by the entire third paragraph of the review document. Appeal decisions are well suited for superimposed schematics because they are highly unstructured, heterogeneous sources of information (e.g., decision and review letters have no physical structure other than a sequence

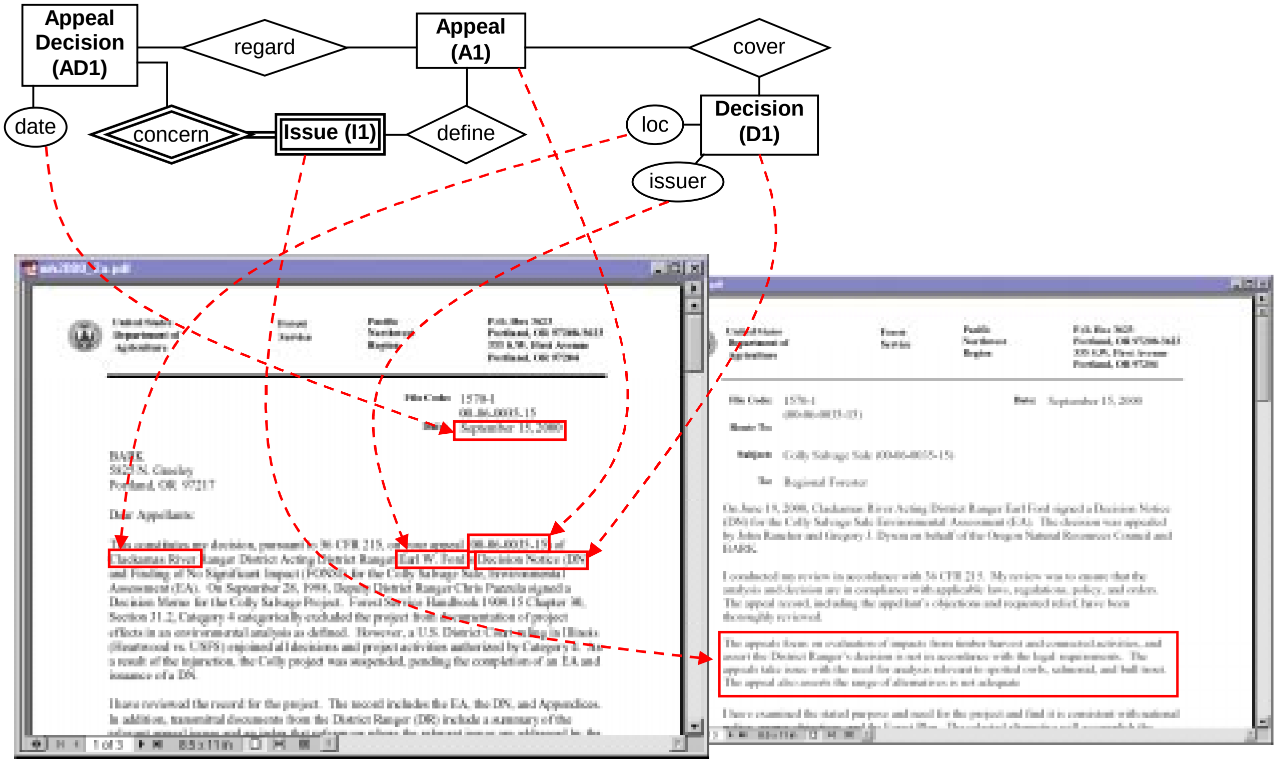


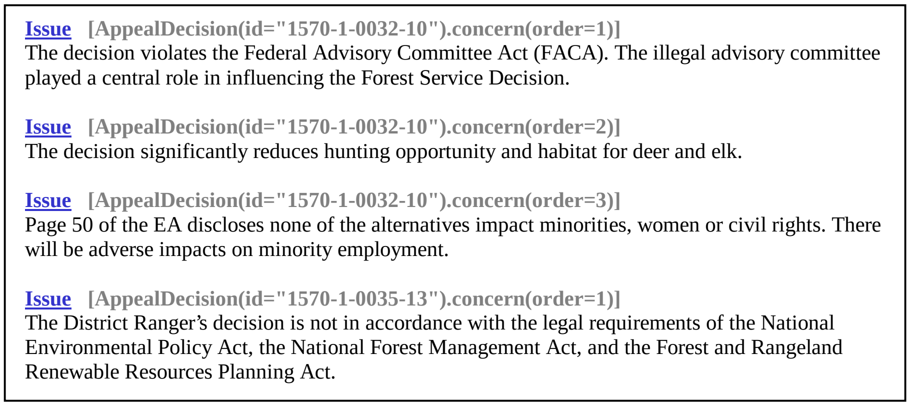
Figure 3: A portion of decision memo (on the left) and accompanying review memo (on the right) shown with a portion of the populated superimposed schematic (at the top).

of letters and words) with important, conceptual content. Schematics provide access to documents without modifying their contents, i.e., they do not add semantic markup directly to the document. Schematics also offer more than simple nested hierarchies of document structure, they contain exactly the concepts of interest represented through an E-R style schema. In general, any given document may have many associated schematics.

We consider document authors as well as schematic designers, schematic populators, and end-users as (potentially) separate people. A schematic populator may create a new schematic instance for each appeal decision allowing natural resource managers (the end-users) to easily browse appeal decisions. For example, a resource manager might begin with an Appeal Decision entity, navigate to the determination(s) and their associated issues, and then browse to see if the issue was raised on behalf of an organization. Schematics serve as richly-structured, conceptual guides for underlying documents. They can be viewed both in and out of place, i.e., viewed without seeing their associated documents or with the documents open to navigate selections in context.

Schematics also enable queries, which can be answered across the collection of all schematic instances. When a natural resource manager is pondering a decision, she might like to know what sort of issues were recently raised for similar decisions. And at a more strategic level, the USDA Forest Service routinely analyzes appeals to track the most important issues and trends, in the mind of the public. These tasks are tedious and labor-intensive, requiring appeal decisions to be read individually. Superimposed schematics are designed for both purposes: information browsing (introducing structure of interest over an unstructured universe of information) and collection-based querying. We see both tasks as being well suited for the Adaptive Management Portal (Delcambre et al., 2001b).

While most documents provide limited addressing capabilities (e.g., page and byte offset, or section,



Issue [AppealDecision(id="1570-1-0032-10").concern(order=1)]
The decision violates the Federal Advisory Committee Act (FACA). The illegal advisory committee played a central role in influencing the Forest Service Decision.

Issue [AppealDecision(id="1570-1-0032-10").concern(order=2)]
The decision significantly reduces hunting opportunity and habitat for deer and elk.

Issue [AppealDecision(id="1570-1-0032-10").concern(order=3)]
Page 50 of the EA discloses none of the alternatives impact minorities, women or civil rights. There will be adverse impacts on minority employment.

Issue [AppealDecision(id="1570-1-0035-13").concern(order=1)]
The District Ranger's decision is not in accordance with the legal requirements of the National Environmental Policy Act, the National Forest Management Act, and the Forest and Rangeland Renewable Resources Planning Act.

Figure 4: A virtual document populated from a schematic query. Each issue is displayed as well as a transparent mark for navigating back to the associated issue entity.

paragraph, sentence hierarchies), superimposed schematics provide *enhanced addressing* over underlying documents, i.e., addressing based on the conceptual information of the schematic to access document content. Enhanced addressing can be used for marking from additional layers of superimposed information. For example, consider the virtual document of Figure 4, which is the (notional) result of the query: "For each appeal decision made in the last two years, what appellant issues were raised for decisions concerning the Mount Hood National Forest." The result is shown as a list of issues, however, each issue is stored in the virtual document as a mark into the schematic along with its associated excerpt. By following a mark (i.e., clicking on "Issue") in the virtual document, the schematic can be further navigated, e.g., to find the issue's determination or the name of the deciding officer. Thus, the virtual document is a separate layer of superimposed information with marks into the associated schematic instance.

## 2 Overview of the Superimposed Schematics Data Model

A superimposed schematic is similar to a standard E-R schema and consists of a set of *superimposed entity types*, which can be associated via *schematic relationship types*. A schematic relationship type is considered part of a schematic when all entity types it associates are within the schematic. Not every relationship type is required to be part of a schematic, i.e., schematic relationship types can span schematics.

Each superimposed schematic can have many *schematic instances*, where each schematic instance consists of *schematic entities*. Each schematic entity conforms to exactly one schematic entity type and differs from E-R entities in that attribute values can contain marks. Similarly, *schematic relationships* serve to associate schematic entities, where each schematic relationship conforms to a schematic relationship type and is considered part of a schematic instance when it connects entities within the same instance. Note that a schematic relationship can connect entities in different instances, but of the same schematic. A schematic entity or relationship can be *anchored* by a mark. An anchor acts as a default "location" for the entity or relationship within an underlying document (see Figure 3). We require that all marks support the *excerpt* function, which returns the associated excerpt of the mark.

Since superimposed schematics integrate marks into an E-R style model, a number of interesting issues arrise, for example:

**Key Constraints** Of interest are equivalence tests when attribute values for keys are mark-valued. Normally, the base layer will provide a Boolean-valued function to test equivalence. However, if no such function exists, we treat a mark's value as its excerpt's value, which implies two entities are equal even if they have different marks but identical excerpts (where at least one of the key values is an excerpt).

| *Address* | *Description* |
|---|---|
| `AppealDecision(id=‘‘1570-1-0032-10’’).concern(order=1)` | An address to an *Issue* entity. |
| `AppealDecision(id=‘‘1570-1-0032-10’’).concern(order=1).desc` | An address to an Issue *desc* attribute. |
| `AppealDecision(id=‘‘1570-1-0032-10’’).response#` | An addres to a *response* relationship. |
| `AppealDecision(id=...).result#[Appeal(num=...).define(order=1).resolve]` | An address to a *result* relationship (identified via its associated *Issue*, which is shown in brackets). |

Table 1: Examples of transparent addresses for the Appeal Decision schematic.

**Authoritative Entities** Keys in superimposed schematics have a subtle consequence when mixed with marks. For example, the Reviewing Officer entity type would normally have a key (e.g., on its name attribute). However, we want to store exactly one mark (for the officer's name) for each appeal decision that the officer reviews. Further, each such "occurrence" of the officer could be spelled slightly differently (e.g., sometimes with a middle initial). We introduce *authoritative entity types* (i.e., keyed entity types that serve as domains for other entity types) to represent key values. Relationships connect entities to their corresponding authorities (via *authoritative relationship types*) to denote identity and enable queries with unique values, e.g., to find all decisions where a particular person served as a Reviewing Officer.

**Schematic Instances** Identifying schematic instances is necessary for navigation and browsing. We introduce *entry points* for this purpose. Any entity type with a key can be an entry point. Each instance of an entry point is required to be in exactly *one* and only *one* schematic instance. We require each schematic to define an entry point, thus, an entry point serves as a schematic instance identifier.

We support two types of marks, *opaque* and *transparent*. Opaque marks contain application specific addresses (e.g., those generated by MS Excel or MS Word), whereas transparent marks contain semantically meaningful addresses (e.g., URLs and XPath/XPointer addresses). Transparent marks can be created and examined out of context, without application intervention, which is not true for opaque marks. For enhanced addressing, we define an addressing scheme (Bowers et al., 2002) for schematics that supports transparent marks. We take a conservative approach, i.e., we restrict addresses to refer to single entities, relationships, or attributes, since the model for schematics does not support arbitrary collections (e.g., schematics only support single valued attributes). Not all schematic items might be uniquely addressable. Intuitively, to be uniquely addressable, an entity must have some form of key or be reachable via a unique path from a keyed entity. This constraint also applies to relationships, since relationships are identified by the entities they associate. Examples of valid transparent addresses over schematics are shown in Table 1.

We require all marks to support the *resolve* operation, which dereferences a mark by opening and (possibly) highlighting the corresponding information selection. For example, for an appeal decision issue, we can call the *resolve* operator on its anchor to view the associated sentence(s) in context. In addition, we treat a schematic instance as another context. Thus for transparent marks into schematic instances, resolve should open the appropriate instance and highlight the selection (i.e., entity, relationship, or attribute).

## 3 Related Work

Superimposed schematics differ from other superimposed models (Biezunski et al., 2000; Delcambre and Maier, 1999; Delcambre et al., 2001a; Phelps and Wilensky, 2000; Lassila and Swick, 1999; Delcambre et al., 1997; Nanard and Nanard, 1993) in that they are highly structured and based on an E-R style data model. Marks are essential for schematics, e.g., although wrappers (Abiteboul et al., 1993; Carey et al., 1995) provide new layers of information, no explicit mechanisms are provided to regain context. With opaque and transparent marks, schematics provide fine-grain addressing for a potentially wide-range of heterogeneous sources, whereas URL-based approaches (DeRose et al., 2001; Lassila and Swick, 1999; Biezunski et al., 2000) have limited sub-document addressing. Even with XPath/XPointer, only XML documents are markable at sub-document granularities. Multivalent Documents (Phelps and Wilensky, 2000) support fine-grain marks, but only fo a specific document model that all other sources must be mapped into. Finally, we have encountered special-purpose schematic-like behavior, e.g., in the Distributed Annotation Server (Dowell et al., 2001), sequence “landmarks” are used to attach annotations to genetic information.

## 4 Future Plans

We have implemented a simple browser for Appeal Decision schematics and plan to extend it to support other Forest Service document types. We also plan to explore simple extraction techniques to help populate schematics, e.g., one could imagine simple tools to harvest marks for deciding and reviewing officer entities. Finally, we plan to develop query capability for schematics. Although considerable work exists on query languages for E-R models, schematics introduce new issues such as: search results that contain marks, expressing mark navigation (for queries ranging in and out of context), and authoritative relationships for identifying entities.

## References

S. Abiteboul, S. Cluet, and T. Milo. Querying and updating the file. In *VLDB*, pages 73–84, 1993.

M. Biezunski, M. Bryan, and S. Newcomb. *Topic Maps*. ISO/IEC 13250, 2000.

S. Bowers, L. Delcambre, and D. Maier. Superimposed schematics: Introducing E-R structure for *in-situ* information selections. Submitted for publication, 2002.

M. J. Carey, L. M. Haas, P. M. Schwarz, M. Arya, W. F. Cody, R. Fagin, M. Flickner, A. Luniewski, W. Niblack, D. Petkovic, J. T. II, J. H. Williams, and E. L. Wimmers. Towards heterogeneous multimedia information systems: The garlic approach. In *RIDE-DOM*, pages 124–131, 1995.

P. Chen. The entity-relationship model—toward a unifed view of data. *ACM TODS*, 1(1):9–36, 1976.

L. Delcambre and D. Maier. Models for superimposed information. In *ER Workshop on the WWW and Conceptual Modeling*, pages 264–280. LNCS 1727, 1999.

L. Delcambre, D. Maier, S. Bowers, M. Weaver, L. Deng, P. Gorman, J. Ash, M. Lavelle, and J. Lyman. Bundles in captivity: An application of superimposed information. In *ICDE*, pages 111–120, 2001a.

L. Delcambre, D. Maier, R. Reddy, and L. Anderson. Structured maps: Modeling explicit semantics over a universe of information. *International Journal on Digital Libraries*, 1(1):20–35, 1997.

L. Delcambre, M. Weaver, T. Tolle, D. Maier, E. Landis, S. Bowers, P. Toccalino, F. Phillips, N. Steckler, C. Palmer, J. Norman, R. Tummala, and S. Varde. Similarity search for harvesting information to sustain our forest. In *(Unpublished Proceedings) DG.O Conference*, pages 155–158, 2001b.

S. DeRose, E. Maler, and D. Orchard. *XML Linking Language (XLink) Version 1.0*. W3C, 2001. Recommendation 27-June-2001.

R. D. Dowell, R. M. Jokerst, A. Day, S. R. Eddy, and L. Stein. The distributed annotation system. *BMC Bioinformatics*, 2(7), 2001.

O. Lassila and R. Swick. *Resource Description Framework (RDF) Model and Syntax Specification*. W3C, 1999. Recommendation, 22-Feb-1999.

J. Nanard and M. Nanard. Should anchors be typed too? an experiment with macweb. In *Hypertext*, pages 51–62, 1993.

T. Phelps and R. Wilensky. Multivalent documents. *CACM*, 43(6):83–90, 2000.