RICE UNIVERSITY

DATASTRUCTURES:
Physical Form for Community Knowledge

by

John Michael Bacus

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

Master of Architecture

Approved, Thesis Committee:

[Signatures]

Christopher Hight, Assistant Professor

Mark Wamble, Visiting Cullinan Professor

Brian Heiss, Visiting Lecturer

David Brown, Assistant Professor

----

Houston, Texas

May 2004
Abstract

Datastructures: Physical Form for Community Knowledge

by John M. Bacus

The shape and function of the contemporary library is approaching a structural crisis, as the rate of publication continually increases and the media of publication diversifies and digitizes. In this thesis, I propose a number of alternatives to these looming problems.

The proposals result from investigations into basic structural properties of information, especially as expressed in academic research and publication, where meaning is largely the result of linking (“this idea furthers this related idea”) and context (as a field of study, a body of work, and institution or research group, etc.)

The proposed library system, called a “datastructure” in a melange of cultural practice, public policy, structural system and technological infrastructure.

Additional research notes (beyond those contained in this document) are kept in a project wiki that can evolve and adapt as the work requires in the future. You can access this wiki at http://think.rice.edu
Acknowledgements

Thanks first and foremost to my excellent thesis committee: thanks to Chris for having an open office where any hare-brained notion was fodder for the best kinds of arguments; thanks to Mark for kicking off my library interest in “The Circuit and the Cell”; thanks to David for keeping my feet at least somewhere near the ground, and thanks to Brian for knowing all about how things work. And a special thanks to Lars for always showing up to point out how ludicrous it could turn out to be. Couldn’t have done it without any of you.

But this work should really be dedicated to my family, who gave me the space, time and support to plow ahead in spite of ever mounting environmental obstacles. To Mom and Dad- thanks for knowing how important this was to finish even when I didn’t; to Meri- thanks for more things than can be numbered, and of course: to Lita- thanks for being just who you are... my mid-semester baby!

Finally, special thanks to @Last Software for giving me the opportunity to be here at Rice for 18 months while we were pushing so hard to make SketchUp even greater than it already is. Also, thanks to the guys at Wolfram Science for pumping my head so full of juicy info at NKS Summer School.
A Special Note on this Publication

In many ways, this project has become a critique of the current paper-based practices of academic publication and of traditional library structures in general. It is not without some irony that the lasting record of this work will be a book - one produced to fit within exactly the structures that I spent such effort reconstructing.

The best record of this work will remain the project wiki, which was constructed and maintained as the primary source of communication throughout the semester of study that concludes with this document. That record can now be found at http://think.rice.edu/

The contents of this publication are culled from the project wiki, and represent only a subset of the complete (and likely ongoing) research on this and related topics. In addition, the wiki remains online into the foreseeable future, and will continue to develop beyond the confines of this book.
Contents

A "Knowledge Structure" 9
Entity-Attribute Relationships 11
Multiway Systems 18
The Desktop 22
Mobile Desktops 28
Master planning 37
Knowledge Contexts 45
The Hardware 50
Walking Shelf simulation 60
A Distributed Library 69
Occupying Fondren 75
Reprogramming Fondren 81
Works Consulted 96
Introduction

In physics, information is known to have material properties. It manifests at high contrast boundary conditions of many different kinds. Through observation and consideration, information can be digested into knowledge, and it can be encoded in persistent forms. Once encoded, knowledge must be shared, reconsidered, edited, commented upon and occasionally discarded as interest wanders through fields of inquiry.

The process of conversion from information to knowledge is dependent on structure. As information is parsed into structure, inter-relationships can be found that link quanta of knowledge into larger and larger structures.

We now refer to this linking process in terms of the World Wide Web where linking has become easy to implement. The "hyper link" is an old concept, however- researchers have always made reference to prior or supporting works. As flows of information increase, however, the density of cross-links also increases.

Consider the modern library as an embodiment of knowledge. It is an archetype now in crisis for (at least) two distinct reasons. First, the yearly increasing rate of publication threatens to overwhelm both the physical space of the library and the curatorial abilities of the librarians who must separate the "signal" from the "noise". Second, "books" are no longer the strict and authoritative embodiment of contemporary research as more and more researchers turn their attention to the much more plastic digital media available to them.

What is the shape of contemporary knowledge? And what is the best physical interface for contribution to it? I propose a new structure- a "datastructure"- to dynamically embody knowledge for our accelerated times.

This datastructure takes many forms and operates at many scales, from the simple arrangement of symbols on a page to the dynamic flow of physical assets through a research community. It is a complex architectural problem that must look beyond the traditional confines of bricks and mortar to consider issues with technology, public policy and cultural practices, and the basic concepts of information and computation.

In the following pages, I will present a number of related provocations inside and around the problem of the contemporary library in an effort to tease out some operable solutions.
In the first section, I look at some structural principles for managing knowledge in the most abstract forms. There are three parts to this. First, I looked at the way in which notes are added to a 'wiki' (such as the one I established to house my ongoing work on this thesis). Second, I looked at the various kinds of "entity-attribute" relationships common in XML schema design. Finally, I looked at multiway systems, which can be used to model the logical expansion of axiomatic systems.

Following these abstract studies, I look at the desktop- the primary physical interface between a researcher and his research information.

Jumping scales, I next apply the geometric systems explored on the desktop to issues of master planning- particularly as related to the Rice University campus plan.

With some notions about what might be relevant at small and large scales, I worked to design an informational system that could organize and protect information, while allowing researchers to study, publish and comment on their work. My approach was a context-based one where information was given additional readability through association with other related works subject to rules of encapsulation at several different scales.

Next, I sought ways to express the physicality of these essentially informational contextual planning diagrams. The result is a series of "hardware" designs for elements of a new kind of library. The pieces represented here are the Hybrid Knowledge Object, the Desk (two versions), the Collaboration Proxy, and a Mobile Bookshelf that is capable of some self-organizing behaviors.

A simulation of this mobile bookshelf learning to "walk" can be found on the thesis wiki, but for this publication, a series of still frames are presented showing several steps in the shelf's evolution.

Finally, I began looking directly at the Fondren Library as a test case for the conclusions I had reached. This testing took two related by functionally different paths, resulting in two separate conclusions... a "reasonable" one, and a "provocative" one. The reasonable conclusion takes the form of a PowerPoint presentation intended to spark more radical rethinking of the future development of the Fondren Library, and it is based on current technologies, public policies and achievable social practices. The second conclusion is one in which I extend my new propositions for library development to their most exaggerated extremes in an effort to fantasize a possible ultimate future development.

This project (and especially this publication) represents only a beginning to a much
longer research project now underway to find effective scalable solutions to the universal informational problems of creation, storage, search and retrieval, and commentary. For more detailed information and to track ongoing developments on this project, please visit http://think.rice.edu
A “Knowledge Structure”

While it is tempting to think of knowledge as timeless, it is in fact dependent to a large degree on causality, and therefore is dependent on time as an organizing principle. In studying the shape of my project wiki, I identified the following datastructure as an efficient model for the creation of knowledge.
In looking at the way in which topics are added to my project wiki, a simple form for representing information became evident.

1. As new topics (notes, thoughts, papers, etc.) are created, they are strung out sequentially like beads on a string.

2. In retrospect, it becomes clear that there are affinities (links) between nodes. These affinities form attractive bonds. In more complex models, different kinds of bonds might be formed (replusors, effectors, etc.)

3. If the nodes are free to move (or constrained partially by the temporal links between them), the string will fold into a knot much in the same way that an amino acid chain folds up to form a protein.

4. Once folded, secondary nodes can be found where several like nodes have contracted into close proximity. These meta-nodes may themselves form links to one another, allowing for a second level folding to occur. As the chain of nodes develops, it might re-link and re-fold in this manner many times.
**Entity-Attribute Relationships**

In XML schema design, information is structured with tags. The character of the tag is somewhat arbitrary, allowing for a number of fairly complex structures to be encoded. The following diagrams explore these general entity-attribute relationships.
a knowledge atom
an Entity with many Attributes
multiple Entities with a common Attribute
an Entity as an Attribute of another Entity
a nested Attribute
a Semantic Web
Multiway Systems

In a multiway system, every possible combination of elements in a list that follow from the application of a set of transformation rules can be found. Multiway systems can be used to model linguistic combinations, such as those created in the process of academic research. Of course, it is comforting to note that even those linguistic systems with rigorously defined rule sets (such as mathematics) cannot, according to Gödel's proof, be complete.
Multiway System:

8 generations in the evolution of a multiway system with simple rules
Multiway System:
8 generations compressed to remove duplicate results
Multiway System:
8 generations arranged in a force-directed diagram
The Desktop

The Desk is a tool for the manipulation, consumption and creation of knowledge objects. Objects on the desk are embodiments of knowledge, and it is through their manipulation that the structured work of research is accomplished.

The desk is the simplest and most essential tool of any research project, and it is at the desk that a rethinking of the processes and facilities of research itself must begin.
**Loading a Book**

To add a book to the Desk, a relationship must be established between the physical artifact of the book and the informational objects to which it is related. Here’s an idea for the UI of this operation.

1. A unique boundary-aware tag must be added to the book if it does not already have one. RFID Tags will soon be a part of typical consumer production practices, but until this is a standard practice, an RFID tag can be added by hand.

2. When the book has been tagged, it is placed on the Desk. The Desk senses the presence of an artifact, but does nothing more than establish a boundary context (and an RDF/XML object) for it.

3. If the researcher wishes to operate on the content of the book (to take notes, have the text read, translated, or to read commentary by other researchers), he strikes the book sharply to “knock loose” the embedded information.

4. This information can be partially assembled by the Desk based on whatever information is available for it. For example, if the book’s RFID tag contains an ISBN number (a likely case if the tag is included by the publisher), bibliographic info, Amazon affinity matches and commentary- and possibly eText, etc. can be retrieved via the internet.

5. The Researcher is free to edit and augment the information contained in this shadow at any point - the Desk merely assembles what it can find without the Researcher’s intervention as starting point. In addition, the book artifact can be removed from the Desk without disrupting operation of the hybrid informational object. A book removed from the Desk is regarded as “hidden” from the Desk’s perspective.
Organizing and Reorganizing the Desk

To understand the desk as a computational system, it is necessary to look objectively at the formal systems in play on the desktop. Research can be understood as the gradual sorting and resorting of information, finding links between ideas and forming new hypotheses based on observations of the patterns these links make.

This process, when observed on the desktop, takes on a distinctly physical character. Sorting information can be seen in the reorganization of texts on the desktop. The book is a kind of informational token, and its placement is critical to understanding the research at hand.

A User Scenario

What follows here is a design for a new kind of user interface for a researcher’s desktop that takes into account the hybridization of digital and physical information assets, networking of researchers into larger research teams, and some notions of commentary and peer-review.

As multiple artifacts are added to the Desk, it takes on a “cluttered” appearance. Books can be piled together (either physically, or using only their shadows). Work in progress may be left “open”. Each artifact on the Desk is surrounded by a “knowledge context” that initially encloses only that artifact.
2. Relationships between objects can be established in two different ways. First, the objects can be linked together with lines into a graph. Second, objects can be enclosed within sets. The dotted lines indicate a set, the arrows indicate a graph. This is the visualization of the basic entity-attribute data structure.

3. Notes can be added to the mix as research continues. Notes are active participants in the system as a whole- they can subscribe to a security infrastructure that allows them to be private or to be shared with collaborators.
4. Simple logical queries can be added to the Desk. For example, in the case of the book in the upper-right corner, the Researcher wishes to have more information about related works gathered. A query to (Amazon) gathers this information automatically, including four suggested additions to the Desk. The researcher can purchase digital copies of these works immediately.

5. In periods of inactivity, the Desk will attempt to organize itself by shuffling around those elements not explicitly pinned in place by the Researcher:
6. Research is seldom done in isolation, and any effective research tool must allow for collaboration. In this model, the desk can be associated with the desks of other researchers working along related paths of inquiry. To make a link across desks, a line is drawn off the edge of the desk.
Mobile Desktops

While thinking further on the ideas in the Desk topic, I developed an alternate notion for a physically mobile desktop. In essence, the desktop is made from a network of elastic bands arranged in a regular tessellation. As objects are added to the network, it dynamically adjusts to a low-energy state.

It is fairly straightforward to model possible behaviors of such a system using the same spring-embedder algorithms found in automatic graph layout software. Here are results from two different tessellations.
Master planning

For a centralized library, it is most practical to locate the facility equidistant from the majority of related facilities. On the Rice University campus, the Fondren Library is conveniently located at the low-energy center of the campus. However, discarding a central library in favor of a network of distributed libraries, a new conception of master planning is required.
dynamic masterplan: current Fondren Library location
dynamic masterplan:
Fondren located algorithmically
dynamic masterplan:
centralizing strategy now in use
known affinity constraint:
a link between known areas of research affinity

interest area:
sized to match the number of researchers active in a given area of study

migration goal:
interest areas are used to provide goals for the autonomous shelving’s migration cycle

dynamic masterplan:
pre-layout, with no facility constraints in place
known affinity constraint: a link between known areas of research affinity

interest area: sized to match the number of researchers active in a given area of study

migration goal: interest areas are used to provide goals for the autonomous shelving's migration cycle

dynamic masterplan: post-layout, no bond to current facilities
known affinity constraint:
a link between known areas of research affinity

facility constraint:
an attraction to the physical structures that house academic processes

interest area:
sized to match the number of researchers active in a given area of study

migration goal:
interest areas are used to provide goals for the autonomous shelving's migration cycle

dynamic masterplan:
pre-layout, with facility constraints in place
known affinity constraint: a link between known areas of research affinity

interest area: sized to match the number of researchers active in a given area of study

facility constraint: an attraction to the physical structures that house academic processes

migration goal: interest areas are used to provide goals for the autonomous shelving’s migration cycle

dynamic masterplan: post-layout, with facility constraints in place
Knowledge Contexts

In large part, the management of knowledge depends on an appreciation of “context”,
where context defines the manner in which an informational asset should be read. In this
section, scalar context is derived up from the scale of the text, to the scale of the desk,
to the research team, to the research consortium. Obviously, there could be larger scales
considered. To some more limited degree, smaller scales could also be considered. Think
of the following diagrams as a kind of informational “Powers of Ten”.

asset address (IPv6): globally unique, context aware, and easily networkable

use policy (CC): Using the Creative Commons system of machine/legal/human readable use restrictions

asset location (URL): a new internet protocol for the desk server

publisher's copyright: digital assets protected against copying by digital watermark policed by publisher as author's proxy

trusted value: Digital signatures and associated "TrustValue" from web-of-trust

author's document context policy: allows author to manage subsequent versions of text, offer annotation and corrections in dynamic way.

access policy (*nix file permissions): visibility and editability tokens to define security policy for the actual information.

author and researcher's context policies: the researcher grants access to collaborators and retains edit privileges
asset address (IPv6): globally unique, context aware, and easily networkable

go-co-orderate: where available, a lat/long coordinate pair for the logical center of this context

asset location (URL): a new internet protocol for the desk server

use policy (CC): Using the Creative Commons system of machine/legal/human readable use restrictions

Trust Value: Digital signatures and associated "Trust Value" from web-of-trust

access policy (*nix file permissions): visibility and editability tokens to define security policy for the actual information.

researcher's desk context policies: the researcher grants access to collaborators and retains edit privileges
asset address (IPv6): globally unique, context aware, and easily networkable

geo-coordinate: where available, a lat/long coordinate pair for the logical center of this context

use policy (CC): Using the Creative Commons system of machine/legal/human readable use restrictions

3ffe:6a88:85a3:98d3:1319:8a2e

research group’s context

Trust Value:
Digital signatures and associated "Trust Value" from web-of-trust

access policy (*nix file permissions): visibility and editability tokens to define security policy for the actual information.

desk://godelStudy@fondren.rice.edu

research group’s context policies:
the research group grants access to collaborators and retains edit privileges
asset address (IPv6): globally unique, context aware, and easily networkable

geo-coordinate: where available, a lat/long coordinate pair for the logical center of this context

use policy (CC): Using the Creative Commons system of machine/legal/human readable use restrictions

research consortium's context policies: the research consortium grants access to collaborators and retains edit privileges

research consortium's context policies:
The Hardware

Over the course of development of this project, several unique pieces of hardware have been designed to accommodate the computational proposals being made. In this section, the primary pieces of hardware are each described in some detail.
encapsulation: The hybrid informational artifact is enclosed in a meta-frame that binds the artifact to its informational shadow.

physical artifact: physical artifacts in the library must be preserved; they can function both as repositories of knowledge, and as phicons for manipulating it.

books stay put: Books, subject to physical gravity, are a way to "pin" mobile environments in place. A book placed over an informational shadow on a worksurface will anchor the shadow in place.

RFID: the entire informational shadow (where possible) or a pointer to the shadow's location on the network is contained in a non-volatile way inside the artifact.

shadows can move: Representations of informational objects are subject to rules for organization; when the system enters its sleep cycle, unattached shadows are free to move. "Unless otherwise inspired, the shadow tries to return to the artifact."

informational shadow: the collected digital assets that have been associated with the artifact through use.

hybrid knowledge object: an encapsulation containing the physical asset with any/all associated informational shadows.
overhead projection: svga or better video projector

RFID in books: each book is tagged with RFID according to hybrid object schema

touch-sensitive tablet surface: opaque Calcomp InterWrite MeetingBoard or equiv.

conventional construction: furniture easily integrates with existing library furnishings

First Generation WorkDesk: achievable using current off-the-shelf technology for a cost of +/- 5,000$
low-friction track:
pleats hinged at one corner to allow easy pan of display material in use

flexible image membrane:
eink or similar large-format high-resolution flexible display material

reservoir of new display material:
display material deployed in bulk

RFID location loop:
in table frame

modular pleat:
pre-folded in simple pleat for expansively deployable worksurface.

Second Generation WorkDesk:
achievable using currently developed (pending implementation) technology
**continuous connection:**
local system is tied to systems at the next context up in scale.

**servo actuators:**
Dynamic framework responds to local repositioning events and sleep-cycle organizing rules.

**processing node:**
Single Board Computer for routing, motion control and proxy service.

**plasma display:**
ruggedized, with pressure-sensitive overlay as from SmartTechnologies (SmartBoard).

**virtual attractor:**
computed affinities for "virtual attractors" which set goals for the mobile processing nodes.

**future growth:**
As required, additional units may be added either at the end of a line, or (more difficultly) in the middle.

**Collaboration Proxies:**
a support infrastructure for remote collaborators.
Mobile Shelving: semi-autonomous, random-access system for storing books and book-equivalents

**continuous line:**
the shelves, like the text and the catalog, are arranged in a continuous line

**processing node:**
upgradable CPU for routing, motion control proxy service, and local cache of digital assets

**random access stacks:**
shelving modules are 1d arrays, allowing random access. Gravity may reconfigure the array in some orientations

**virtual attractor:**
computed affinities for "virtual attractors" which set goals for the mobile processing nodes

**servo actuators:**
Dynamic framework responds to local repositioning events and sleep-cycle organizing rules

**local addition:**
As required, additional units may be added either at the end of a line, or (more difficulty) in the middle.
a datastructure: this model contains 336 modules, in two separate (but intertwining) strands
Fondren Library, 2nd Floor: transitional form; 45% of traditional shelves removed
Walking Shelf simulation

Executed in the simulation environment “Breve”, this simulation suggests some kinetic strategies a walking bookshelf might adopt as it learns to navigate its physical and informational contexts. The following stills show steps in the movement of the simulation in both free and constrained environments.


A Distributed Library

The distributed library will have a form- a new kind of form- that is in constant motion. As inquiry wanders across the campus, the library will follow it. It is impossible to predict the precise form at any moment- just as it is impossible to freeze it long enough for a complete analysis.
**Occupying Fondren**

As the stacks convert from static shelving to walking shelves, the collection might go through a variety of different hybrid states. The walking shelves, as they replace traditional storage, infiltrate and re-occupy space as it becomes available. Finally, all existing shelving has would be replaced, leaving a constantly shifting kinetic landscape of ad-hoc research.
Reprogramming Fondren

Many of the proposals made in this project are provocations intended to expose problematic aspects of contemporary research space. However, there are a number of simple steps that could be undertaken with currently available technology, political and social practices. The following section is derived from a PowerPoint presentation titled "Reprogramming Fondren" that outlines a proposal for research library reform.
Some Open Problems for Libraries

1. The global rate of print publication is increasing by approximately 5% each year. At the same time, digital publication via the internet is growing at a rate of almost 12tb per month.

2. There is an uneasy relationship between physical and digital assets; they represent two competing and only occasionally supportive cultures of information.

3. No library can be authoritative on all subjects; inter-library loan is becoming increasingly critical as institutions are forced to specialize.

4. Catalog and search systems are unable to handle increasingly dynamic and diverse informational assets.

5. The “digital revolution” has met resistance in the print community. Copyright issues have lead to prohibitive digital access requirements, and consequently acquisition of printed matter remains a priority.

6. Increasing rate of publication and interdisciplinarity makes editorial decisions increasingly difficult. The “authority” of a digital text is often difficult to establish.

Typical Responses

The problems associated with academic library development are broad enough that a recognizable pattern of typical responses can be seen across institutions that are working proactively to solve their growing problems.

1. Libraries must focus on the content of their collections, not on the medium in which the content is stored. Archival tendencies where the role of the librarian is concerned primarily with the protection against loss or damage of physical assets must change.

2. Copyright issues must be addressed, and several prior initiatives (eBooks, for example) have largely failed. Without a comprehensive re-thinking of digital copyright practices, it will be financially impractical for libraries to acquire digital assets.

3. Once a critical mass of information has been collected, it is access (not ownership) that matters the most to researchers. The library of the future is a gateway to distributed information, not a repository of physical books.

4. Increasing reliance on inter-library loan demands broadening of library consortia, and rethinking of the means of distribution of shared assets. No library can satisfy the needs of all researchers all the time.

Issues Specific to the Fondren Library
The rate of the growth of the Fondren's collection has outpaced the capacity of library systems, staff and facility. At the same time, the size and scope of the collection is judged by many to be inadequate for the needs of a major teaching and research university.

The Fondren Library is ranked 104th place in terms of library holdings (near last place) among all major research libraries in the US. 70% of books in all disciplines have been checked out in the last 15 years, indicating that the holdings are heavily used. 21% of surveyed Rice faculty report holdings are inadequate to support the needs of their undergraduate students.

There is little integration between digital and physical assets in the collection. Library users, particularly those who have grown up with access to the internet, increasingly turn to the internet first for research, where information is much more easily accessed.

Finally, the Fondren Library suffers from regional isolation; it cannot rely on neighboring institutions to supplement the collection. The closest major research libraries to which Fondren researchers might travel to find resources not in the collection are several hours away by car. It is critical, therefore, that the Fondren Library maintain a diverse collection suited to the needs of a very wide variety of needs.

**Insufficient Space for Print Collection**

Accepted standards for library shelf space suggest that shelves should be considered “full” at 85% capacity. Fondren stacks are at 100% capacity on two floors, and at 95% on the others. The Fondren Library houses over 2 million volumes, 2.6 million microforms and 14,300 current serials and publications. 250,000 volumes (14% of collection) are housed off-site in closed stacks.

The collection has already been aggressively weeded; 11,000 volumes were removed in 1994, and an additional 13,000 were removed in 1995. Currently, approximately 1,500 volumes are weeded annually.

Yet the rate of acquisition is approximately 50,000 volumes per year, for a net increase of 48,500 volumes per year. Space intended for ancillary services (study/research rooms, etc.) has been lost to the ever growing stacks. Existing stacks have been compressed together as much as physically possible, resulting in cramped, ill-lit and difficult to use spaces.

**The Library Service Center**

Opening soon, the Library Service Center will provide space for 1.75 million volumes
(or "book equivalents"). It will provide a place to store less-used but still important to preserve assets separate from the main stacks. The LSC will be a closed stack facility, open by appointment only with 24 hours advance notice given. It is five miles from the main campus.

The Library Service Center is a short-term solution intended to solve immediate problems, and to relieve pressure from the main facility while more concrete plans are formulated.

**Insufficient Workspace for Researchers**

Study carrels and research offices are inadequate to the needs of individual researchers, and are difficult to for researchers to acquire.

There are insufficient group study spaces, and they are inadequate to the needs of extended research projects.

Spaces that allow the integration of multiple asset media types- or which provide network access- are practically nonexistent.

**Awkward Asset Search/ Retrieve**

SIRSI's “WebCat” system does not provide adequate search functionality.

False positive results are common.

There is little integration with other library systems- even those inside the Fondren's library consortia.

When an asset has been selected via WebCat, it is difficult to find it in the stacks because they have become so fragmented.

There is little usable integration between the catalog and the physical layout of the stacks.

**Intellectual Property Issues**

The ownership of creative work remains a major blockage in the drive towards digitization in every creative field.

The battles over Napster, and the continuing efforts of the RIAA and MPAA are evidence of the depth to which this is a problem.

Progress is, however, being made in the music industry through projects like the iTunes Music Store. It appears that Apple has found an acceptable DRM policy, and has been able to implement it in a way attractive to consumers.
On the other hand, eBook initiatives have largely failed. Digitization projects focus on works not subject to copyright (Project Gutenberg, for example), or on relatively esoteric (thought internally complete) projects that can be sold as a whole.

Amazon has taken an important step forward with “Search Inside”, though industry response has been guarded.

An acceptable DRM policy for text must be found.

**Creative Commons**

An understanding of the problems of intellectual property must begin with an understanding of the motivations of various types of authors.

While a writer who depends on profits generated by her writing understandably adopts a protective stance, not all writers have this motivation.

“Creative Commons” copyright follows precedents set in the open source community (the GPL, etc.) to allow for copyright that considers:

- Attribution
- Commercial use
- Derivation
- Requirement to share

**A New Vision: the “Library Function”**

Information is found at boundary conditions. As Gregory Bateson said, “Information is the difference that makes a difference.”

Standards are ways to make public space. Just as with language, two individuals must agree on a standard in order to communicate.

The critical functionality of a library has to do with a cycle of storage, retrieval, and publication. It is a support structure for the process of thought and the creation of knowledge, which remain uniquely human practices.

The “library function” is a distributed network of interconnected storage and retrieval objects. This may coincide with an institutions goals, or with goals of authors, but it is not dependent on either of them.

The “library function” is inherently spatial, though it requires an architectural approach that includes bricks and mortar only as one of many constituent elements.
Goals for a New Fondren

Expedite access to authoritative research information

Local caching

Informational shadows

Context-relative search and navigation

Support a community of commentary

Encourage informal publication

Incentivize commentary

“Trust” function

Foster interdisciplinary team work

Inter-disciplinary reading rooms

“Matchmaker” function

Solve space issues for print collection

Engineer systems where knowledge is an emergent property

Interventions

Implement a new distributed, random access catalog. Discard the hierarchical catalog, and allow assets to reorganize by context, relation and proximity.

Distribute the collection into public reading rooms across the campus so that research groups are closer to the assets they are using.

Give researchers with specific research projects the responsibility for the assets relevant to their work by extending the collection to the desktop.

Digitize everything possible, and tag the originals with RFID to broaden distribution of information while allowing preservation of the artifact.

Distribute responsibility for the editing and sorting of the collection to those who know the subjects the best.

Close the main stacks to the public once a critical mass of information has been distributed. Compress the stacks into an archival form, and reclaim public space for work and study and collaboration.

Design, build and maintain physical structures for the above.
Proposed Work

Establish or join broader library consortia and expedite inter-library loan systems. Find acceptable business solutions to Fondren’s “net-borrow” position.

Adopt/Establish standards for the storage and distribution of digital assets. Develop a hybrid object standard compatible with RDF that allows physical artifacts to carry and project an informational shadow.

Deploy an RFID infrastructure capable of locating assets anywhere in a distributed network of contexts. A careful balance of asset security and researcher privacy must be maintained.

Implement “trust” function (a machine-readable extension of the Honor Code) and deploy it to manage a community of commentary and casual publication.

Deploy affinity-based master plan for interdisciplinary reading rooms

Deploy Hosted Reading Room and Desk objects.

Reclaim Fondren facility for interdisciplinary works paces

Library Consortia

Fondren must continue to develop lending relationships with other major research libraries.

There is no longer a requirement that member libraries be geographically close to one another. A shipment from New York does not cost more than a shipment from Austin. Nor does it take any more time.

Digitization makes possible an unprecedented sharing of collections without risk of loss or damage- and without need to remove a book from local circulation.

Consider a closer relationship with Amazon. Bring the commercial world inside the library consortium. Libraries can be more than simple consumers of information.

Quick turnaround for asset acquisition

Pre-existing digitization infrastructure

Accepted system for community commentary

Academic validation and provision of expert commentary to Amazon

Expedite Inter-Library Loan
Because of the size and scope of Fondren’s collection, alliance with larger library consortia will lead Fondren into a “net-borrow” position.

Adoption of book scanning services at Fondren may provide sufficient value-add to other members of a larger library consortium to offset the net-borrow issue. Books loaned to Fondren will be returned with a scanned copy attached.

Transportation of physical assets is problematic. Assets may be lost, or damaged- and shipping costs can be prohibitive. Fondren should help to enable a digital-only asset sharing practice for its library consortia members.

Key to this is the establishment and implementation of a hybrid object standard that links physical assets to informational shadows within a manageable DRM framework.

**Asset Digitization**

Benefits of digitization:

- Creation of asset “backup” to hedge against loss/damage to original
- Generation of a “machine-readable” copy that can be translated into alternate media
- Machine-readable text is searchable, increasing speed and accuracy of citation.
- Digital assets are highly portable, and can be shared with affiliate libraries efficiently. Also-

  a digital asset can be shared by multiple researchers with ease.
- Tagging systems enable the addition of commentary to a text.
- Digital assets effectively take no space to store.

Detriments:

- Prohibitive copyright practices make sharing of digital assets problematic.
- Without a comprehensive security/versioning policy, it can be difficult to establish the “authenticity” of a digital asset.
- Duplication distances authors from their work, which can be misrepresented or altered without their permission. The medium is too plastic.

**Costs of Digitization**

Outsourced domestically, a book costs approximately $.50 per page ($150.00 for a 300 page book) to digitize, with a 48-hour turn around. Offshore, costs can be reduced $1-4.00 per book, though with a sacrifice of speed and accuracy. Also- the easiest path to
scanning requires the destruction of the scanned book.

New book scanning robots (like those available from 4DigitalBooks) promise to reduce the cost of scanning to a negligible amount, with a 10-20 minute turn around. In addition, the scanned book is not damaged in the process.

When the costs associated with scanning a book approach the cost of round-trip shipping, scanning will become desirable in the process of inter-library loan. This may provide the path of least resistance for the adoption of wide-spread book scanning.

**A new Informational Standard**

A "hybrid object" combining physical and digital attributes

- RDF compatible XML
- Add unique serial number for specific asset
- Embed in RFID tag affixed to asset
- Build the asset's "informational shadow"
- An eText standard for proper formatting
- UTF-16 for character encoding (covers all known human symbolic languages)
- MathML for specific mathematical symbology
- CSS to provide rich text capability
- Follow W3C standards
- Dynamic location encoding
- A dynamic Gödel-type enumeration

\[
\text{assetID} \ast \text{containerID} \ast \ldots \ast \text{ContainerIDn} = \text{very large but unique number}
\]

**Bibliographic and Citation Information**

- Call-numbers in all known cataloging enumerations (LOC, etc.)
- ISBN number
- A pre-formatted "full citation"
- Unique assetID
- Links to: (other documents that cite this document)
Asset Location
Degree of certainty that the asset is located where it appears to be
“Last seen with” information if asset has been removed from the network
Availability and status of asset’s informational shadow
Dynamic location encoding
A dynamic Gödel-type enumeration
assetID * containerID1 * ... * ContainerIDn = very large but unique number
Context-relative directions can be decoded from this enumeration

Commentary
Free-form commentary from previous asset users
Follow “trust function” principles for visibility of commentary
Encourages interdisciplinary by identifying the “authority” of a text in the context of research currently ongoing or resident within the institution
Identifies “experts” within the university community that a researcher may wish to consult for more information
Provides valuable perspective on the relative merit of the work in consideration, and provides alternate directions for future research.

Informational Shadow
Machine-readable text and scanned images as they are acquired.
Commentary, as following institutional web-of-trust
Machine manipulatable (search, translate, reformat, etc.)
May be accumulated slowly, or only as needed- removing the burden of mass digitization.

Deploy an RFID Infrastructure
RFID tokens (proximity cards) to all library users containing:
Borrower’s privilege policy
List of assets for which the borrower is responsible
The private key in a strong public-key encryption key pair
RFID readers at all security points
Register entry/exit events, and allow them only if assets are paired with an appropriate borrower privilege policy
Automate checkout/checkin of assets to a responsible party
RFID Readers in Reading Rooms and Works paces
Generate dynamic catalog of room contents, searchable across Fondren network
RFID Tags in all library assets
Programmed and added to assets as they leave traditional library circulation
Tag implements new “hybrid object” data standard

Leaving the Fondren Network
Policy should encourage assets to remain inside the Fondren network at all times, though assets may travel dynamically to arbitrary geographic locations while still remaining “within the network”.
Assets kept within the Fondren network would be keyed to location, not to borrower. Privacy could be achieved by accessing a resource without removing it from the system.
“Checkout” would of course be possible, and would have the effect of attaching the borrower’s name to the asset as the person of last-known responsibility.

Expedite Community Publication
Implement “trust function” to manage editorial functions
Publications are made available in digital form on the network.
DRM wrappers are applied according to the preferences applied by the holder-of-copyright.
“Authoritative” copy of work remains in the control of the author
Fondren acts as a primary cache for public access.
Foster a “community of Commentary”
Space is provided in hybrid objects to house relevant commentary
Relevance (and authority) of commentary is managed by “trust function”
Commentary increments the commentator's trust value, and therefore incentivizes researchers to contribute.

Amazon provides a model for this function, but it is too informal. There is no way to judge the commentator's legitimacy as a critic of the work in question, and not all comments are worth reading. Slashdot's "karma system" may be a better model.

**Authority (the “trust function”)**

A graph structure with researchers as vertices, and attributions as edges

Vertices are assigned numeric weights as an indication of the “trust” associated researcher's authority on a subject.

Commentary (Y) on a work is the product of a rating integer (-10 < X > 10) and the commentator's own trust value

Y = Xa

The trust value (T) of a published work is the sum of all commentary (Y)

T = \( \sum [Y_1 + ... + Y_n] \)

Publication (and commentary is considered publication) adds to the researcher's trust value in the form

a = a + T

Slashdot's "karma" system is an example of a web-of-trust in action.

**Distributed Collection**

The new Fondren Network will be composed primarily of an interconnected network of temporary reading rooms.

Assets in the current Fondren Library will remain there until they are "checked out". Once checked out from the old catalog, they may be checked back in only when no longer considered relevant to current research.

Assets checked out from the old Fondren catalog are placed by the researcher in a convenient Reading Room, where they stay until needed by another researcher.

Books may travel from Reading Room to Reading Room as research interests vary.

Researchers searching for a particular asset may find that the Reading Room in which it is found will contain many other interesting assets that were previously unknown.
Reading Rooms provide pre-sorted and dynamically relevant micro catalogs of specifically important assets.

**Dynamic Master plan for Collaboration**

A dynamic (though algorithmic) affinity-based system for locating reading rooms at the midpoint between collaborators.

Using a “matchmaker function”, the master plan dynamically allocates spatial resources for collaborative workspace.

University faculty interests (as documented in university publications) are pattern-matched to provide a basis for intentional (or unintentional) affinity.

Resulting “affinity lens” objects are arranged spatially on the university campus. When a location has been determined, a mobile reading room can be deployed.

In some cases, the “affinity lens” may come to rest over an existing facility. In this case, provisions are made to adaptively reuse the facility.

**Hosted Reading Rooms**

Schools, Departments and Research Groups will be invited to set aside space in their facilities for a Fondren Reading Room.

Reading Rooms house ad-hoc collections of Fondren resources relevant to proximate research.

Assets can be added to the Reading Room by any Fondren user, and can also be removed by any Fondren user.

Assets seen to be rapidly moving from Room to Room will be added to an acquisition list (or optionally leased from Amazon) until the moving assets stop moving.

The Reading Room records entry and exit events (via RFID) for assets.

These spaces must remain open to the research public according to current Fondren access rules.

Proximity cards (RFID tokens) may be used to control access for these Reading Rooms, but it is suggested that the host provide regular staff for the room where possible.

**Reading Room Components**

Horizontal work surfaces with comfortable seating sufficient to allow collaboration in small work groups.
Informal meeting space for casual meetings

Storage space for small (but targeted) collection of Fondren assets

Coffee or beer... and snacks

A/V module

Asset Messaging System

Assets are found by researchers through the Asset Messaging System

Context-relative, not global; assets only message one context up or down in their dynamic call number:

Call-and-response: The Fondren network is polled for a particular entity. If the entity is in-network, it responds with a “here I am:” message.

Signaling mechanism attached to RFID tag affixed to asset; for example, a low-power electroluminescent “pummer”

Enables context-relative navigation

“I’m in the Fondren Network”

“I’m in the Anderson Hall Reading Room”

“I’m on the third shelf from the left”

“I’m the book that’s blinking”

Reclaim Fondren

As the stacks are gradually distributed across the new Fondren network, space will become available for new uses in the Fondren Library building.

The library as an institution seeks to foster and develop research work for students and faculty at Rice University. Unburdened of the print collection, the institution is free to pursue new projects.

Space for research offices, individual and group works paces, and interdisciplinary meeting rooms can be established.

The Fondren can focus on problems of data manipulation, translation, and distribution as a support for new interdisciplinary research on campus.

Finally, the Fondren can focus on standards generation and maintenance to assure that informational assets remain both viable and accessible in perpetuity,
The Proposal in Summary

Overlay the existing LOC catalog with a new non hierarchical catalog of library assets organized by context, relation and proximity.

Build a new public asset identification standard that allows asset hybridization (digital/physical), aggregation and authoritative commentary.

Deploy a network of ad-hoc reading rooms arranged spatially by an affinity-based “matchmaker function”. Encourage ad-hoc sorting of physical library assets into these rooms.

A work surface
Storage space for physical assets
Space for discussion
A dynamic master plan
“Contextual” wayfinding system

Foster a digital community of authoritative commentary by establishing a machine-readable “trust function” leveraging existing digital signature technologies

Deploy an RFID infrastructure capable of locating hybrid physical/digital assets within this new distributed network of contexts

Reclaim open space for inter-disciplinary collaboration in the library as assets are sorted out into the reading rooms.

Closing Thoughts

Is the centralization of the Fondren Library an asset?

Is the friction in interdisciplinary work an asset?

How about the problems of publication?

Generally speaking, are the inefficiencies of our current system unappreciated contributors?
Works Consulted

The following references are the primary works consulted in the course of this research. However, a large number of additional resources were consulted in various lesser ways as well. For a complete record, with commentary, please consult the project wiki at:

http://think.rice.edu


Hockney, David ed. by Nikos Stangos. 1993. That's the way I see it. London: Thames and Hudson.


