November 15, 2004

University of California Libraries

Kudos

- Evolving collections and service environments that add value to education and life-long learning at a variety of levels – both individual projects and the nsdl as a whole
- Rich policy environment
  - Which seems actually to influence practice (with regard to collections and metadata)
- A testbed for applied digital library research that has taken us at least as far as NSF DLI 1 and 2.

Kudos 2 – SUPERLATIVE WORK ON

  - DL architectures (an early entrée into HE of a layered service model) – other examples JISC DNER, OCLC, RLG, shared services at UC
  - OAI harvesting
  - More good and focused work on metadata than exists anywhere at all
  - Evolving understandings of user needs, of measuring use, and of sustainablity
  - Evolving understanding of online instruction – needs, values, and opportunities
- You have a brilliant browse-level interface
- You have radiated influence and raised the benchmark for digital library development by encouraging good practice (Roy's example)
- And you have not pinned your successes solely on shaping individuals' behaviors (work on recombination augmentation augers well for…

Kudos

Obviously there is a great deal more to be done

Sustain the current high-level of effort while doing more to

- Demonstrate power of recombinance and the value of layered approach to service delivery – unpacking silos with MD enrichment
- Dramatically enrich the content base
  - Tackle the web, selectively – NDIIPP, Internet Archive, etc
  - User-driven gap analysis – Am South and Am West
  - Do more, perhaps, to create incentives for and lower barriers to contributions of good collection content
- Vigorously extend the open-source tool base (to do all of the above)
- And do all of this on declining NSF funding. And it is here, naturally, that I want to focus my remarks, on the imperatives of sustainability planning

Some business planning questions?

Is the NSDL supposed to sustain both the level and the variety of investment currently represented in the NSF's NSDL portfolio? Is it intended to sustain the sum total of :

- the core integration service
- collection efforts (the primary assets as well as the portals)
- research efforts
- service and patheway efforts?

If an NSDL.org was making the investments (rather than the NSF) and the investment horizon was concentrated entirely or largely on sustainability and business objectives, would the investments be made in the same way?

- I can't answer the question. I can say, that if the NSDL.org is intended economically to sustain all of the affiliated projects, it will not easily succeed
- Much more practical would be an NSDL that offers a limited, clearly focused, and self-sustaining set of products

But here's the conundrum

- and you'll know it better than me… The NSDL is a complex and highly interdependent enterprise in which the success of any one aspect requires success in other areas
  - Individual collections will likely succeed best where they are seen in the context of, perhaps even integrated with other like collections. They rely on the core integration service and on the work of those projects that add value by utilizing content in the MD database to build rich services
  - The core integration service's success requires contributions to the MD Repository and repository utilization in the form of value-added services built using its APIs
  - The value added services that may be supplied, for exapmle, by the pathways, will rely upon content availability in the MDR and the core integration service
- So does NSDL.org stumble right out of the starting gate because in order to succeed it has to take on and return the sum-total of NSF's current investment that it is unlikely immediately to match?
- I don't think so

To focus our thinking, let's make some limiting assumptions. Based on some of NSDL's own user needs research, let's ask what services it might offer that could actually generate revenue

Introduce CDL work for NSDL (science), Am West (humanities and social sciences), and for UC shared services (broad discipline but U based and including core collection, area studies

Results in two areas: on users' expectations (form the environment in which we work) and on their needs (shapes the work that we do)

First: users' expectations

Where educational information is concerned users expect:

- The speed and simplicity of the search engines (Google)
- The convenience of e-commerce (Amazon)
- The reliability, authority, and integrity of [library/academic or professoinal society/state curriculum]
- They are willing to pay for value-added service; less so for the "infrastruture" that supports them
- A long way to go in the educational space before individuals will "pay" for anything (the Questia quandary)
- High degree of reticence about paying for government funded goods and services

Second, some users' needs - 1

Curated collections matter (a lot) - relevance, reliability, integrity, brand

- *K-12 and community college teachers* and media specialists under huge time constraints. "Less interested in 'doing research' than in quickly locating a few useful, high-quality teaching materials'. They want to go to authoritative and reliable collections where materials are selected to support teaching or learning in particular disciplines and at particular levels. Particularly interested in class-room ready materials, lesson plans, and pathways through them
- *Academic and public librarians* want curated collections because they extend the holdings they offer patrons. More catholic in their interests because their user base is broader. But they want and need to be clear about what the collections are and who they are for. Enthusiastic about American West, less so about NSDL because they do not yet understand the NSDL's collections aims or audiences.
- *Undergraduate students* want curated reliable, selective, and authoritative collections that bear on a particular information need. Where such collections exist, they gravitate to them and away from Google when completing assignments
- *Graduate students* consume information like locusts so curated collections are of less interest though they want to know basically what the boundaries are.
- *Faculty* are like graduates but more conservative in finding new trusted online information sources. They tend to get pointed at these by their graduates.

Some users' needs 2

Interactivity is important and adds value – annotation, export, specialized analysis

- C*ommunity college faculty and K-12 teachers* and media specialists want classroom ready materials that can be slotted directly into a particular class, preferably with a localized narrative web that makes them immediately relevant. Although there is considerable anxiety about who is responsible for creating these materials and the narrative web that surround them , it is clear that tools that enable their construction are immensely desirable. Particular significant are tools that enable users to select, annotate, and organize, materials that are found within a collection, and to save selected materials for future use, possibly exporting them into other software environments (e.g. powerpoint, locally maintained web pages, etc.).
- *Academic and public libraries* are interested in federating selected relevant virtual collections with local holdings, and surrounding them with context, look, feel, and functionality that meet local users' needs
- *Graduate students and faculty* require specific suites of tools to capture and manage citations (as appropriate for developing footnotes and bibliographies), and to create and manage their own personal reflections (research notes) on the materials they review. Where they are required to build teaching materials for online use, the tool suite they require extends accordingly to include annotation, export, etc.

Some users' needs 3

- Content, even large portals that organize access across a fair breadth of it is considered part of the infrastructure for which users are reticent to pay. This is particularly the case with content created through investment by federal funding agencies.

Summary of research

- Digital libraries add value to information by organizing it into coherent, highly differentiated collections that are developed to support particular modes of inquiry and/or user communities. Users access these collections with the current generation of sophisticated search tools, but also through a range of highly specialized end-user services.
- Google, by contrast, adds value to information by amassing it into vast, largely undifferentiated collections which users examine with increasingly sophisticated search tools.

Based on this review, the single most powerful value proposition

- Curated, special, service-rich collections that target specific communities with appropriate look, feel, functionality, service and substance - examples
- If this is true, then revenue generation is going to happen at the front end (of service) not the back end of technology, infrastructure, or even content development
- If this is true, then NSDL.org has to generate demand (and revenues) not from amongst end users or even end user communities but amongst the service providers that serve those users and communities, mediating between them and the Natl Sci Dig Library

Following this logic

- NSDL could focus its efforts on a small number of defined rich information – targeted by discipline and user community and buliding out and controlling the entire stack from the content, through the infrastrucutre and tools, and critically, to the value-added service front end
- Less ambitious and less interesting then the current mission (all disciplines all levels), this approach is destined to disenfranchise and to turn the NSDL into yet another publisher which adds value by controlling the entire scholarly communication chain from the IPR in the content to the delivery front ent
- The approach fails to utilize the power of the layered service model

Another approach…

- Adheres to the original ambition and builds on strength (always useful in sustainability planning to be something you are)
- Here, NSDL is envisaged not so much as a publisher but as a broker that mediates relationship between content providers and service providers and that does so in a way that facilitates developments on either side of that equation

According to this approach

NSDL.org would bet the farm on utility infrastructure that would enable third parties that carry the appropriate domain knowledge, brand recognition, and credibility with specific user communities to build these service-rich collections. Those utilities would include:

- Recombinant, quality content, much of it uncooked (since cooked content is the revenue generating stuff)
- Infrastructure, tools, and incentives that encourage content contribution to the MDR, and that substantially lowers the costs involved to service providers in using this content to build these high-quality service-rich and targeted collections. A lot of those tools already in place MDR, ….)
- 1, 2, 3, or more exemplars that demonstrate possibility and iteratively address key technical and business issues. Eat our own dog food (or dog's breakfast)

Lets build some use cases to put some tangible flesh on this skeletal abstraction

- A textbook publisher supplements its textbook with lesson plans based on NSDL content. Lesson plans are freely available but more meaningful when used with a core text that is commercially available
- An educational publisher creates course-packs that use the best of the NSDL, incorporating it into a narrative context of its authors' own creation – again access to underlying content is free
- A scientific society wraps the discipline with educational materials and a "best of the discipline" newsletter that makes annotated reference to openly accessible content in the NSDL and other openly accessible repositories of scholarly and educational information
- An instructional technology group at a well-known state-wide community college is funded by the statewide system to develop selected learning materials appropriate for systemwide use and leveraging NSDL resources

These use cases aren't far fetched

- They are built on real-life examples of activity that is already happening and beginning to demonstrate that the value a "publisher" brings into the educational space is not derived from the content that they own but from the way in which they present it
- It is a very powerful and compelling model for a variety of reasons
- Suggests that a sustainable economics for scholarly communication is a little more nuanced than "information wants to be free" even while it locates information content as part of an openly available infrastructure

So a couple of questions worth pursuing

- The first is… what would NSDL charge service providers for?

My answer is this

NSDL would not charge for access to content in the MDR (remains forever open access – though content providers might derive revenues as royalties from third-party publications based in part on their content)

Nor would it charge for access to the basic tool suites (hopefully open source) that service providers use selectively to gather, organize, and present that content (any more than it would charge for use of its APIs

Leveraging its own infrastructure and expertise, it would charge service providers for

- Consulting (how do I do it?... Who could I do it with?)
- Key data services
  - Metadata analysis and enrichment (taxonomies)
  - Content aggregation (web crawling) and management
  - Bespoke collection development
  - Bespoke service and tools development
  - Digital preservation
- It could also leverage content and technologies, and the value added and focused services built with them to create its own breed of new service offerings. Thereby stimulating further investment in new areas while meeting unmet needs (fobbing off services as cdl?)

There's a third question and it is a big one and it has to do with the cost of creating the elemental content that populates the NSDL MDR

- At least, the model I've set out above assumes limited financial flow down from the front-end value-added service to the underlying content provider.
- It also assumes that the revenues available as royalties are unlikely to cover the entire cost of content creation necessary to populate the MDR with rich and compelling stuff

Moving content from the proprietary front end into the infrastructure

- NSDL incentive
  - Adding path of opportunity to the range of current incentives
- Selective NSDL investment (in catalytic, gap filling, and synergistic activities)
- NSDL framework of tools, collection and sustainability-planning guidelines and materials, and evaluation criteria
- NSF incentive
  - Requiring open-access to research publications (including data) resulting from NSF-funded research – not so far fetched; NIH trying the same
- Institutional incentives (implementation of appropriate load balancing, and promotion and reward structures
  - The trickiest of all areas but surely NSF/NSDL can help to create incentives and route maps for change

Organizational desiderata

- Able to make and execute timely business decisions about priorities, processes, protocols, products, and pricing
- Able to invest in support of business objectives
- Trusted and independent
- Able to motivate, create incentives for, facilitate, and leverage co-investment from content providers and service providers
- Able effectively to represent all facets of the NSDL to all comers
- Accountable to shareholders
- Offering transparent value to stakeholders (enhancing, empowering, enabling)

Organizational and governing models

- Highly centralized –operational service will impede flexibility, nimbleness, etc. and may threaten community culture
- Highly decentralized – strengthens community culture but unlikely to execute efficiently or quickly or to escape from the shadow of the operational data service
- Light-weight, service oriented broker or managing agent. Least known but most compelling

Startup funding possibilities

- Membership
  - pay-to-play (ARL)
  - venture fund (DLF)
- Insufficient capital, hard to define membership attributes
- Large entities likely to benefit most (pay less now or more later)
    - Universities
    - State DoEs
    - Scientific research establishment
    - Service providers?