A Proposed Model for Access Integration[1]

SOPAG – August 28, 2002

ABSTRACT

The large and increasing number of portals[2] used to locate information resources managed by libraries and other research collections poses a serious problem for our users. How do they find all the appropriate portals and where should they start searching? To address this challenge, the UC Libraries will have to develop a shared, conceptual vision of how these portals should be integrated to provide efficient and effective access across each campus, as well as at the UC library-wide level.

This paper begins by identifying the challenges inherent in trying to integrate access to library materials in the current environment. It then proposes an Access Integration Model (AIM) that could be used as a framework to understand and address these challenges. The goal of the AIM is to help us decide what portals are needed and how they should be integrated together, in support of users with varying levels of information needs. It is hoped that this paper will generate a UC-wide access integration discussion that will lead to a shared vision and framework to meet the challenges ahead.

THE CHALLENGE

The UC Libraries collect materials that span the whole range of human knowledge. Our Libraries accomplish this by acquiring materials in specific subject areas, languages and geographic areas. These collections are built and crafted to best serve individual campus academic programs. Just as important, our librarians craft ***access*** to these materials to ensure they can be discovered, and therefore used by the constituencies they serve.

Traditionally, the library catalog has been thought of as the primary discovery tool, at least for books and journal titles. Other portals have been employed to complement the catalog, such as abstract and indexing databases for journal articles, and finding aids for archival

---

[1] This paper is based on previous work by the UC library-wide Access Integration Task Force, whose final report can be found at http://www.slp.ucop.edu/sopag/TFAIReport1.pdf and on discussions with numerous other library colleagues. B. Hurley, editor

[2] The terms portal, access system and discovery tool are used interchangeably in this paper.

collections. More recently, technology has allowed libraries to provide additional portals for: governmental numeric datasets (Counting California); electronic pre-print (E-Scholarship); GIS enriched data (ECAI); digital finding aids (OAC); electronic journal titles (CDL Directory); electronic reference materials (CDL Reference Shelf); Internet resources (e.g., INFOMINE, Sage); visual resources; etc.

Discovery tools can be categorized by their importance in broad subject areas. For example: the sciences have tools that provide access to journal articles; the humanities have tools that provide access to books, etc. However, this generalization is by no means absolute. While science librarians may see periodical indexes as their most important discovery tools, they still need to provide access to books through the traditional library catalog.

The large and increasing number of portals, and their varying levels of importance to different academic disciplines pose a serious problem for library users – where do they start? At one extreme, libraries are addressing these problems by creating customized, subject-based websites designed to help users in a particular topic area find the discovery tools that best fit their needs. At the other extreme, libraries are exploring the "scholars portal" approach, which would provide access to a large range of materials via a single portal (i.e. like a Google search engine for academic materials).

The UC Libraries do not yet have a shared conceptual vision of how our many portals should be integrated to provide efficient and effective access for our end-users. Therefore, it is difficult for our libraries to create integrated solutions that work across each campus, as well as at the UC library-wide level. The following sections of this paper describe an Access Integration Model that proposes a framework to address the issues listed above.

THE ACCESS INTEGRATION MODEL (AIM)

The AIM as a Reference Model

The AIM is a ***reference model*** for access integration. Its purpose is to stimulate a discussion on what types of access systems we need and how they should be integrated. Here the term "reference" has nothing to do with traditional library reference services. In the literature, reference models are used to describe the functional and metadata requirements of a system, ***without*** saying how the system should be implemented. Reference models are helpful in that they focus discussion on what a system should do and what metadata is needed to support that functionality.

Using a reference model also allows us to defer implementation decisions that could complicate a discussion on functionality, such as what technology should be used, and how our existing systems (Melvyl, SearchLight, etc.) fit into the picture. If some version of the AIM reference model is adopted, the question then becomes, "how do we modify our existing systems, and/or develop new ones, that together create a AIM-compliant implementation?"

AIM Goals and Assumptions

The AIM is based on an assumption that it is not possible to build a single access system that "does everything for everybody." Current practice bears this out, as we see many systems being developed to meet different needs. In the case of UC Libraries, we have Melvyl, the CDL Directory, Searchlight, the OAC, campus catalogs and databases, as well as many other websites to access e-journals, abstract & indexing databases, government information, etc.

Instead of a single system, the AIM proposes a network of cooperating access systems, each with its own goals, tasks and responsibilities.

The model also recognizes that people work at different levels of sophistication, based on their current "information needs." Keyword searching to find a few good books on dinosaurs may fill a relatively simple need. More complex needs may require special access systems that allow for sophisticated searching in highly area-specific metadata indexes. For example, a researcher might need to use a paleontology database to find specimens of a femur bone belonging to a certain kingdom, phylum, class, order, family, genus and species of dinosaur.

The goal of the AIM is to create a framework to help us decide which access systems are needed and how they should be integrated together, in support of users with varying levels of information needs. At the highest level, the access and integration goals are supported by three types of portals, which are the:

- ***Global Access Portal***, which provides a "scholar's portal" or "Google-like" service that uses keyword searching and reports backs search results grouped by material type (e.g., 1,200 books, 35 maps, 120 images, 12 numeric datasets).

- ***Subject Portal*** that is a version of the Global Access Portal, which reports back search results grouped by material type in predefined subject area.

- ***Material Type Portal***, which provides more sophisticated search and display services for their specific materials type, than can the other two portals.

Put simply, there are two tasks that must be addressed in designing a service based on the AIM. ***Material Types*** that are meaningful to our users (e.g., book, images, maps, fossils) must be assigned to our content, as they will be used to group search results. We must also identify the different portals needed to support discovery and display services for these material types. That is, we need to decide in advance what Material Type Portals we require (e.g., books, maps, EADs, GIS) and when they should provide information to the Global Access and Subject Portals.

At this point, working with the AIM may seem like an academic exercise. However, designing an AIM compliant system would take into account current realities. For example, our current metadata has material types that we can use as a starting point. There are also existing access systems created by many different communities, which can be used as Material Type Portals.

Understanding AIM Goals and Portals

The ***Global Access Portal*** is designed to provide services that mimic the best features of Internet search engines, such as Google. The general idea is that a keyword search will return a list of responses that can be used to drill-down to more detailed levels of granularity.

Internet Search Engines respond with a single category of results – web pages. However, the Global Access Portal needs to present many categories of materials. Therefore, its first level of response will be grouped by material type (e.g., 1,000 books; 4 maps, 110 visual materials). The reasons for this approach are:

- It will prevent categories with a smaller number of responses (e.g., the 4 maps) from getting lost in the overall search result set.
- Many fields that can be used to define "material type" already exist in our metadata (e.g., MARC Format and Dublin Core Type elements; data file types, such as TIFF and .WAV). Using existing metadata, when possible, to define material types means that we do not have to bear the cost of creating new metadata.
- Carefully defining material types will minimize the problem of having the same object appear in multiple search result groupings, which may be confusing to a user. Carefully predefining types, such as books and maps, would help to ensure that objects are classified under a single material type.
- Finally, and most importantly, grouping search results by material type is an approach that our users will understand.

The Global Access Portal may be too general a tool for users with complex information needs. Therefore, the ***Material Type Portals*** will have more sophisticated user interfaces and use search indexes designed for specific categories of materials. For example, a visual materials portal may have a graphic-based user interface that displays image thumbnails in response to a search. Or, a structured-text portal (e.g., texts encoded with TEI) would have a textual interface. Note that the more general global access and subject portals will not be able to display all materials types, but could call a Material Type Portal to provide this functionality. For example, clicking on a search response in the Global Access or Subject Portal for a GIS map could launch a Material Type Portal that specialized in the display, navigation and manipulation of GIS maps.

The ***Subject Portal*** will integrate best with web-based services crafted by librarians to serve their constituencies. The challenge here is to ensure that the correct level of subject classification metadata exists. In most cases one would expect the subject classifications to be fairly general (e.g. engineering, music, etc.). However, the model allows for detailed subject categories. The cost of creating the detailed subject metadata may be a limiting factor. It also should be noted that an item with a specific material type could have multiple subjects, and therefore, be included in different Subject Portals.

Portal Example

The following wire-diagrams are provided to help understand the portals. The material types used in these examples are illustrative only. Let's start with a user entering a search on "Napa Valley."

GLOBAL ACCESS PORTAL

Your search for "Napa Valley" found 138 Items

72 Books
30 Websites
21 Maps
5 Journal Titles
8 Manuscripts
1 Image
1 Video

The user clicks on the “Maps” link and gets….

GLOBAL ACCESS PORTAL

Your search for “Napa Valley” found 21 Maps

1. Napa, American Canyon and vicinity / Cartographic Department,
California Automobile Association.<2000>
Author: California State Automobile Association. Cartographic Dept.
Published: San Francisco : California State Automobile Association, c2000.
Holdings: Earth Sci G4364.N2 2000; .C3; Case B

2. Napa, American Canyon and vicinity / Cartographic Department,
California Automobile Association.<1999>
Author: California State Automobile Association. Cartographic Dept.
Published: San Francisco : California State Automobile Association, c1999.
Holdings: Earth Sci G4364.N2 1999; .C3; Case B

3. Cycler's road map of part of the Sacramento Valley and Vicinity
[including Colusa, Yolo, Napa, Butte, Yuba, Sutter, Solano and Sacramento counties,
[Geo. W. Blum]<1896>
Author: Blum, George W.
Published: [San Francisco : The Author, 1896, c1895]

Finally, the user clicks on the link in #3 and a Material Type Portal is called to display the image (try it, the link should work). In this case, the Material Type Portal is Insight, from Luna Imaging, Inc.

How the AIM Approach Differs from a Traditional Library Catalog.

An AIM compliant system is conceptually very different from a traditional library catalog. The library catalog is a megalithic access system that provides discovery and display services to traditional library material types (e.g., books, journal titles, maps, etc.).

An AIM based implementation is a network of access systems working together to provide various integrated “views” of materials, at different levels of sophistication. Material Type Portals offer sophisticated discovery, display and data manipulation services, based on the material type they are supporting. The Global Access and Subject Portals provide broad, general services and draw the metadata needed to support these from metadata catalogs provided by other systems.

The AIM is also better suited to work with non-library materials, which may have their own Material Type Portal, but still can provide appropriate metadata to the Global Access and Subject Portals to support keyword searching. For example, adding support for fossil metadata (e.g., kingdom, phylum, class, order, family, genus and species) to a traditional library catalog would be a major undertaking. In fact, this approach will not scale when one looks at all the different type of metadata schemes in use. In the AIM, the Material Type Portal would provide sophisticated access services. Broad access for users with simpler information needs would be offered through the General Access and Subject Portals.

A TECHNICAL DESCRIPTION OF THE AIM

This section describes the AIM from a technical perspective, which is important for system developers or anyone else who wants to understand the model in more detail. This section can be skipped if one is not interested in the technical detail.

Figure 1 provides an overview of the AIM. From the bottom up, we see the:

1) ***Data Content Layer***, which contains objects with predefined material types (MTx) held in a digital library repository, or on the Internet;
2) ***Material Type Aggregation Layer*** that has Metadata catalogs (MCx) created to aggregate metadata for a specific material type;
3) ***Material Type Integration Layer*** that integrates separate metadata catalogs from the layer below; and
4) ***The Access Integration Layer*** that contains the portals.

Access Integration Model

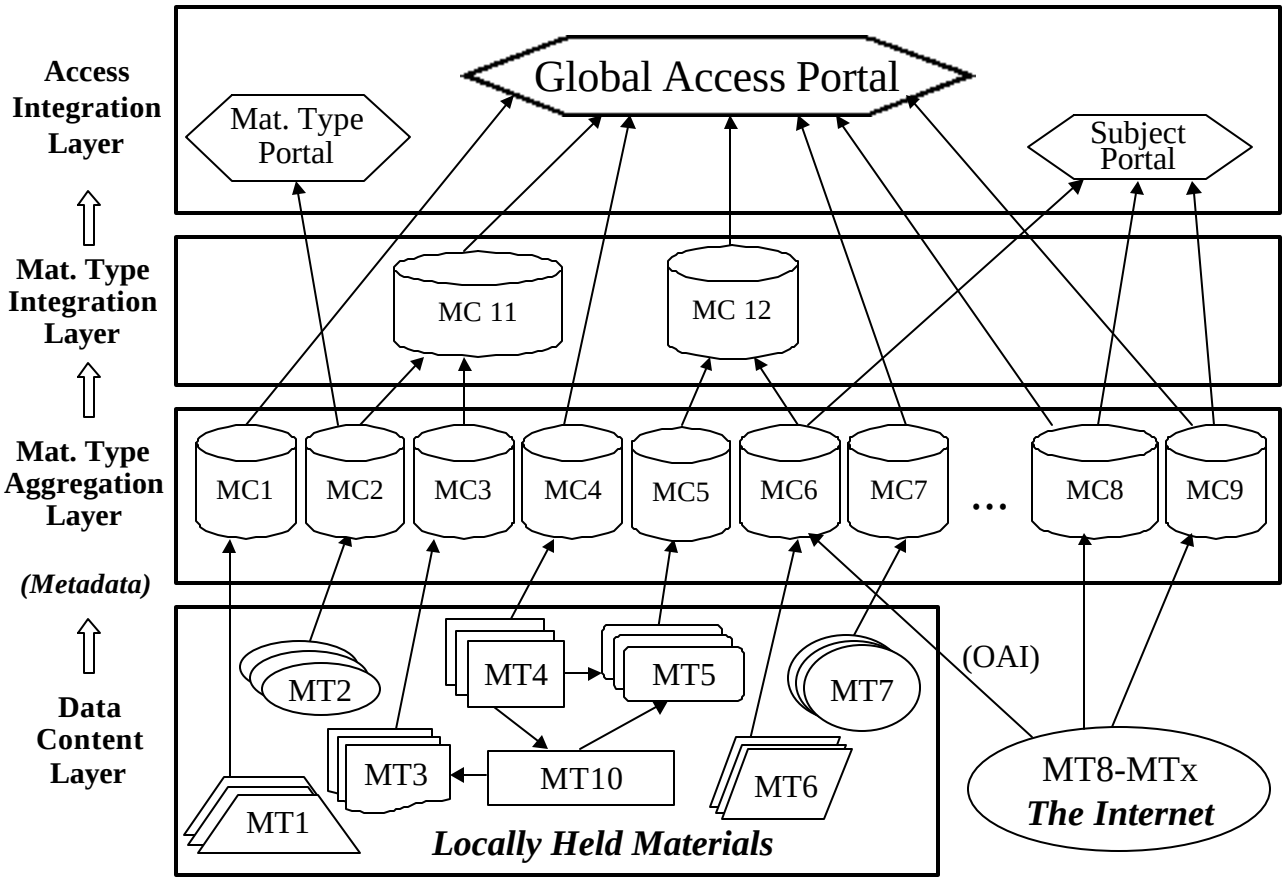

Figure 1

Material Types

The AIM does not name the material types to be used to build metadata catalogs. However, they must be pre-defined by an organization that is implementing an AIM compliant system. Figure 2 is an illustrative example of one way in which material types could be assigned.

It should be noted that material type does not necessarily have a direct relationship to the MARC format type (e.g., monograph, serial, map, etc.), file format (e.g., TIFF, PDF, etc.), or any other existing format definitions. However, using existing format metadata defrays the cost of adding new material type designations to existing metadata.

It is also important to understand that material type is assigned to describe the nature of the item (e.g., book, map), not the metadata scheme used in its description. That is, one would probably not assign material types of MARC, Dublin Core, VRA, etc.

Illustrative Material Types in the Access Integration Model

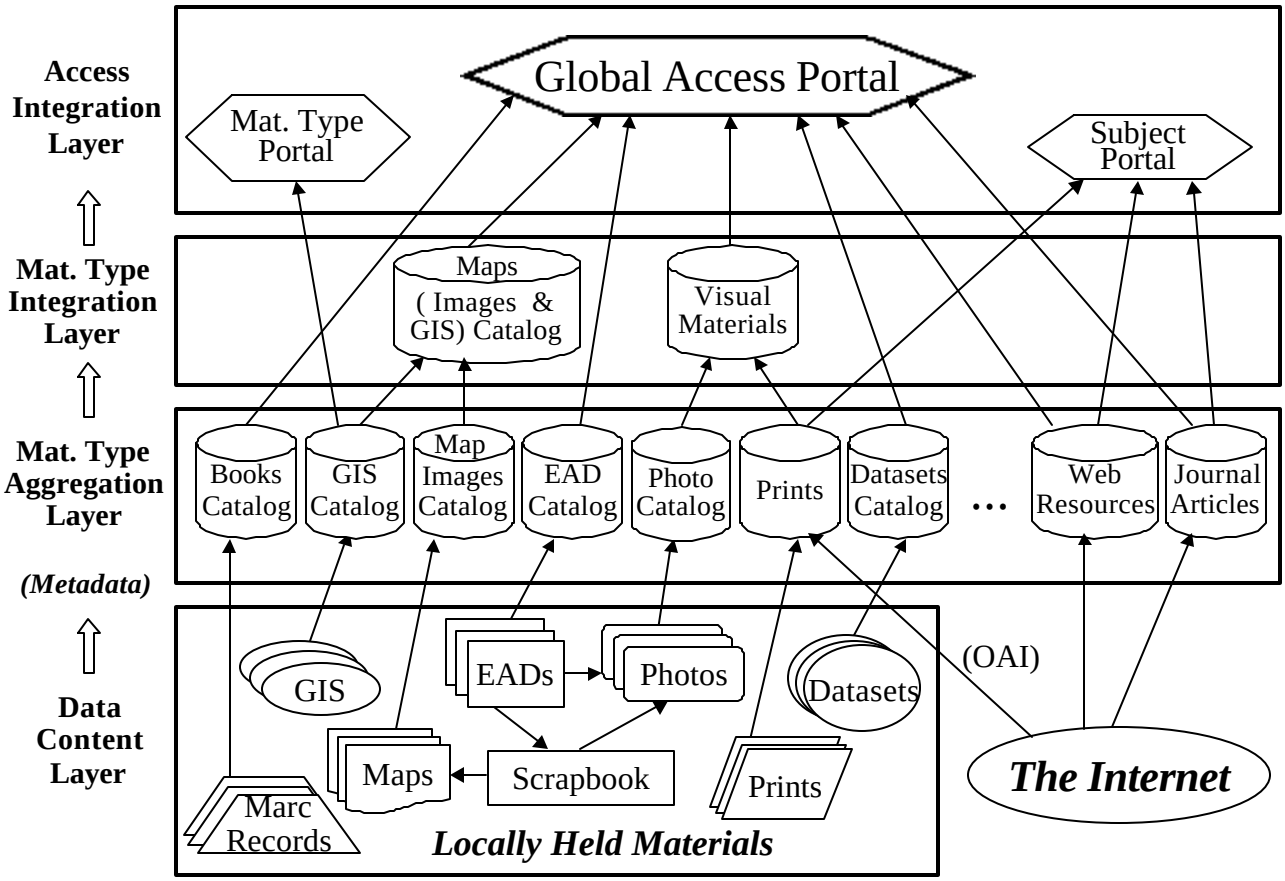

Figure 2

The Layers of the AIM

The Access Integration Reference Model contains four layers. From the bottom-up they are:

The Data Content Layer

This layer holds or points to content, as well as the metadata that describes the content. Content can be held in digital library repositories or on the Internet. Note that physical content (e.g., printed books) can be represented in a digital library repository through its metadata.

The Material Type Aggregation Layer

Here one finds a separate ***metadata catalog*** for each material type. The function these catalogs are to provide the metadata necessary to support searching for a specific material type in the Global Access, Subject and Material Type Portals.

As mentioned earlier, material types are not related to metadata schemes (e.g., MARC, Dublin Core, VRA, etc.). Under the AIM, one would probably not choose to build a metadata catalog of "all MARC records." Instead, MARC records would be spread over the metadata catalogs, as appropriate. For example, the MARC records representing photographs would go into the "photograph metadata catalog" along with entries for other metadata schemes (e.g., Dublin Core, VRA) that represent photographs.

As the AIM is a reference model, the concept of a metadata catalog is actually a virtual construct, rather that an implementation guideline. While the AIM does not specify how metadata catalog services should be implemented, it may be useful to provide examples to help clarify the distinction between the concept and multiple implementation options. A metadata catalog could be implemented as:

- ***A Collection of Metadata Records*** that the Global Access and Subject Portals can use to build their own keyword search indices.
- ***A Simple Access System***, which can report keyword search results and the location of digital objects (e.g., URL, URN) back to the Global Access and Subject Portals in response to a broadcast search. For simple objects like single images, the portals can use the location to display the object in the browser, upon request. This type of simple access system would not be a Material Type Portal and therefore, would not be directly available to end-users.
- ***A Material Type Portal*** that can report search results back to the Global Access and Subject Portals, as well as provide more sophisticated search services directly to end users. The Global Access and Subject Portals may also call the Material Type Portal to display complex digital objects. For example, GIS files, numeric datasets, objects made of multiple parts (e.g. a multi-page digitized scrapbook with envelopes pasted on the pages, some of which hold letters, audio CDs, etc.)
- ***Variations on the Above***. Note that within a single AIM compliant system, metadata catalogs can be implemented in a variety of ways.

The Material Type Integration Layer

This layer allows for the integration of separate metadata catalogs into a new "multi-material type" metadata catalog, or ***multi-material catalog*** for short. Multi-material catalog search results are presented as one grouping in a Global Access Portal result set. In principle, multi-material catalogs do not have to be built from metadata catalogs found at the level below. If there is no special use for separate metadata catalogs, such as a Material Type Portal, the multi-material catalog can be built directly from the data content layer.

An example may help illustrate the use of multi-material catalogs found at the Material Type Integration Layer. As illustrated in Figure 2, we could have two material types called GIS and Map Images. The GIS files have a Material Type Portal that is used for sophisticated searching and rendering (i.e., displaying the GIS files on screen, as a "map"). The Map Images are JPEG files represented in a separate Map Images Metadata Catalog. Under this

arrangement, the GIS and map images search results would be reported up to the Global Access Portal, as ***two*** different categories. By creating a "Maps" multi-material catalog, the Global Access Portal reports only the one category of Maps, in which the search results for GIS files and map images are interfiled. If a user clicked on a GIS file search result, the Global Access Portal would have the GIS Material Type Portal to render the map. Clicking on a map image would result in the Global Access Portal displaying the single map image in the browser.

The Access Integration Layer

As explained earlier, this layer provides three types of access portals. The first is a ***Material Type Portal*** that is specifically designed to provide direct access to one of the underlying metadata catalogs. The second portal type is the ***Subject Portal*** that integrates access across selected metadata catalogs for a given subject. Note that this is only possible if the participating metadata catalogs also hold appropriate subject classification metadata. Finally there is the ***Global Access Portal*** that can be thought of as a "Scholars Portal" or "academic Google" service that integrates access across all metadata catalogs (e.g., books, maps, visual materials).

THE AIM AND INTERNET CONTENT

Integrating Internet content into the Global Access and Subject Portals poses a most serious challenge for libraries, as we have no control over the metadata behind the websites. The following discussion proposes some methods for integrating this content and actions we may pursue under the AIM framework.

In general, Web based resources would be entered into a Web Resources Metadata Catalog[3]. Web based materials that matched a predefined material type (e.g., photos) would be most likely be added to the appropriate material type metadata catalog, as opposed to a Web Resources Catalog. In many cases the entries for the Web resources will have to be manually entered in to its metadata catalog, as libraries have no access to the existing metadata. However, the Open Archives Initiative[4] (OAI) holds out promise that at least some of this metadata can be automatically harvested by program.

Journal article integration is a special case that may represent the biggest challenge for libraries. Currently, many publishers and periodical indexing organizations create the article citation metadata used for searching. Therefore, this metadata is spread out all over the Internet on many different websites (i.e., metadata catalogs). Since libraries do not have control over this metadata, they cannot aggregate it in ways that serve their users. For example, by having a single Global Access Portal search that reports all appropriate journal articles, regardless of the journal in which it was published. Therefore, our portals will most likely be designed to treat journal articles as a special case – perhaps by routing the user to the most likely e-journal website.

However, adopting the AIM would give direction to UC librarians in future negotiations with e-journal publishers, regarding services needed to support the AIM. Libraries could be pressing publishers to provide the "metadata catalog" service that would integrate journal articles

---

[3] If you skipped the technical section, a *metadata catalog* provides the metadata necessary to support searching in three portals for a specific material type.

[4] The Open Archive Initiative (OAI) protocols allows participating organizations to post and/or harvest descriptive metadata from each other via the Internet. For more information, see: http://www.openarchives.org/

into the Global Access and Subject Portals. For example, there is an organization called CrossRef[5], with 149 publisher members, which provides "a collaborative reference linking service -- through which a researcher can click on a reference citation in a journal and immediately access the cited article." The CrossRef database represents "over 6,844 journals with over 5.2 million article records." Libraries could be negotiating with CrossRef to develop a metadata catalog service that returned all journal articles for a search to our portals, regardless of the publisher.

SUMMARY

The goal of the AIM is to create a framework to help us decide what access systems are needed and how they should be integrated together in support of users with varying levels of information needs. The AIM defines the Global Access Portal and Subject Portal, which provide basic keyword searching and then group search results by material type. This grouping was selected because it: is understandable to our users; prevents categories with fewer search results from becoming lost within large result sets; minimizes the need to de-duplicate results sets; and relies on material type metadata that is already widely available. The Material Type Portal provides advanced search, display and data manipulation services for users with more sophisticated information needs.

[5] For more information on CrossRef, see: http://www.crossref.org/