

DLF Service Framework for Digital Libraries

A progress report for the DLF Steering Committee 17 May 2005

Prepared by **Lorcan Dempsey** and **Brian Lavoie** on behalf of a working group comprising Peter Brantley (California Digital Library), Lorcan Dempsey (OCLC), Dale Flecker (Harvard), Brian Lavoie (OCLC), Krisellen Maloney (University of Arizona), Andy Powell (JISC/UKOLN, University of Bath), and Mackenzie Smith (MIT)

1	Context.....	1
2	Why is a Digital Library Service Framework needed?	2
3	Some benefits of a Service Framework for DLF and research libraries.....	4
4	What has the DLF Service Framework Group done so far?	5
5	Some open issues	5
6	Recommendations moving forward.....	6
7	Appendix A: The group's progress to date	9
8	Appendix B: Model of business requirements and services	11
9	Appendix C: Business requirements, business processes, and business functions in practice: the OAIS reference model	13
10	Appendix D: Related work.....	14

1 Context

This is a brief report outlining the progress made by the DLF Service Framework Working Group (henceforth the SF Group). Pending discussion by the DLF Steering Committee, it will form the basis for a wider publication.

The Research Library is undergoing foundational change. Recent years have seen major advances in understanding of service needs and systems to support them. However, the research library community has not yet transitioned to a shared understanding of how a library and its services are organized in an increasingly networked environment.

Think of recent discussion and development in the following areas:

- Repositories and collection management.
- Metasearch and portals.
- Integrating library services with course management systems.
- Exposing library collections to search engines.
- Building services on top of federated collections (see Aquifer, for example).

Although some common responses are emerging, we do not have a shared sense of the specific services we expect in each case, and how they should be presented in different environments.

The absence of common models undermines our ability to develop and design systems efficiently, to create large-scale collaborative activities, and to communicate the value of libraries to other communities. It makes it less likely that library services will be routinely embedded in system-mediated research and learning workflows (the course management system, for example). It slows the development of common or third-party systems which would reduce the costs involved in redundant development activity across many institutions. And it makes it impossible for the library community to mobilize its collective resources to respond promptly and efficiently to changing needs.

This was the context in which the SF Group was asked by the DLF Steering Committee to develop a 'Service Framework'.

2 Why is a Digital Library Service Framework needed?

A Service Framework is a tool with which a community collectively organizes its attention. Typically it provides a pattern which can organize discussion, design or resources. Here, we want to use it as a tool to organize our collective attention to library services in a changing environment.

It is a tool for library directors who are thinking about resource allocation and strategic direction. It is a tool for library staff who are building systems and services. It is a tool for funding bodies and other library organizations who are prioritizing funding allocations. It is a tool for related communities who wish to understand the touching points between their environment and the library environment.

An example of a framework is the *OAIS Reference Model*. This has helped organize some of the attention given to digital preservation activities, allowing progress to be made on the basis of some shared understanding of a problem space.

We need to be clear that a Framework is not a substitute for local innovation, for bottom-up creativity, or for new thinking. It needs to change and adapt.

To further explore this idea we need to look more closely at **services**.

What is a service? A service is any functional component which it is useful to talk about as a unit. At one level, of course, the library itself is a service. The reference desk is a service. The catalog is a service. We recognize that a growing proportion of library services will be encapsulated in automated systems.

A service may be directly delivered to a user interface. However, not all services will be directly used by humans. Some services will be used by other applications, or 'intermediate consumers', which build applications from one or more services. In this case, a service may provide a machine interface to the consuming application. Examples of intermediate consumers might be a campus portal, a course management system, a metasearch engine. Increasingly, library applications themselves will be consumers of services, assembling their functionality from different components.

This modularization will become more common as libraries and their digital environments become more complex and interconnected. They need to be able to build systems flexibly and responsively. They need to reduce the cost of making changes. All of this points to the need for a modular approach, building up applications from simpler parts. This in turn raises several questions about the nature of those components and how they communicate:

- At what levels of granularity and aggregation should services be designed?
- How do we move existing services into a digital networked environment?
- How do we design services to be flexible and responsive?
- What machine and human interfaces should services support?
- How should services be designed so that they are reusable and interoperable?
- What functionality should be split out into a shared service infrastructure? rather than being created again and again (authentication, for example)?

A Service Framework helps us address these questions. Specifically it provides:

- A shared, consistent vocabulary for discussing library processes and services.
- Common ground for talking about processes and services within and across libraries.
- Common ground for talking to other domains – e.g., e-learning or e-research – about how they might take advantage of library services;
- A basis for identifying reusable components and interoperability needs.
- A roadmap for identifying priorities for shared attention or development, and for identifying gaps and 'pain-points'.

We return to these points in the next section. It should be clear from this discussion that the issues are not only technical. In order to develop an effective framework, one needs to understand the 'business goals' that drive service development, and also the type of data we need to run systems effectively. It became clear to the Group that there were two related emphases:

1. To understand and communicate the 'business processes' needed to support library goals and how they are changing as library goals develop.
2. To support the design and construction of interoperable systems to implement those processes.

There is a close relationship between these, especially as more library processes are network mediated. The first is particularly important if the value of libraries is to be understood; the second is particularly important if the value of libraries is to be realized.

The SF Group is convinced that developing such a framework is critical to the DLF if it is to fulfill its mandate, and is vital to DLF members as they address the challenges of the current environment. We are not alone in reaching this conclusion, as we discuss further below.

3 Some benefits of a Service Framework for DLF and research libraries

Here, we expand slightly on some of the benefits a Service Framework would provide. These relate to specific current concerns.

- **A modular approach.** As in industry more widely, we are seeing a general trend towards a more modular approach to system and service development. The reasons are obvious: a modular approach facilitates the flexibility and responsiveness required in an environment in flux. It also potentially reduces the long term management costs of services. It is difficult to think of a modular approach without some sense of how these modules fit together to meet goals. This points us to a 'framework' or architecture. Indeed, without such a perspective it is difficult to see how libraries can easily engage in scalable collaborations within or across institutions. Without a framework, how does one know what to federate and how?
- **Providing a foundation for collaboration among libraries.** Libraries need to strengthen their collaborative efforts if they are to remain a driving force in the increasingly competitive information marketplace. A framework that provides common definitions and shared views of services and processes will facilitate collaboration. Without such a perspective, it will be difficult for libraries to engage in scalable collaborations within or across institutions.
- **Lowering development and ownership costs.** If the library community does not routinize the process of designing and building services and systems, the cost of doing so will become increasingly high, as resources are wasted in repetitive planning and design efforts. There is duplication of effort and reinvention of wheels. In addition, the library community will never realize the economies associated with off-the-shelf solutions, because services and systems will be custom-built and of limited applicability. In short, if we do not have 'patterns' to guide development, then every development project becomes a new adventure – even if another institution has already experienced that adventure. A framework should reduce long-term costs by providing a context within which the consensus necessary for shared or third-party development of tools and services can be established.
- **Expressing the value of the library to other communities.** We have observed that communication with other domains such as e-learning and e-research is problematic in the absence of a shared sense of library services. There is a growing perception that potential library contributions are undervalued or overlooked. Indeed, it is often left to others to define the library contribution, letting caricature flow into the vacuum (for example, the view that the library is little more than a warehouse of books).
- **Aligning library planning.** We are at a stage where it is vital to bring thinking about the 'business' of the library much closer to thinking about the 'systems' that will realize that business. A major and growing part of the library will be manifest through its systems environment; it becomes increasingly important to have good ways of aligning communications between library management and library systems developers, and between discussion about the 'business' of the library and the 'services' which realize it. A

framework is a good way of mediating between 'business' and 'technical' perspectives.

- **Identification of gaps and development needs.** The current environment calls more than ever for libraries to mobilize their collective resources. However, without an overall framework within which to look at things it is often difficult to identify gaps and priorities. Systemic investment and planning is reactive and opportunistic. Furthermore, it may not be clear yet to many in the library community that we are looking at a major new model of resource creation and management, as well as the need to deal programmatically with new 'business entities' which have not been schematized in standard ways and hence are costly to manage (for example, services, collections, rights, licenses). The DLF ERMI work has made a valuable contribution in moving forward one area. However, we need to be sure that we are alert to other development needs.

4 What has the DLF Service Framework Group done so far?

The Group's work to date has dealt with articulating the need for and anticipated uses of a Service Framework for digital libraries; reviewing related efforts and settling a plan of action for moving forward with development of a framework; and developing a model of business requirements, business processes, and services which relates the business of libraries to the services needed to carry out that business. This model will underpin the development of the Service Framework.

Appendix A contains a more detailed description of the Group's progress. The model relating business requirements to services is described in Appendix B. Given its familiarity, the OASIS Reference Model is presented in Appendix C in the terms that we have developed. Related efforts are discussed briefly in Appendix D.

Some more specific work has been done to test the model. In particular we have explored some modeling of two 'business requirements': the discovery to delivery chain, and collection management. This work will be built on moving forward.

The model of business requirements and services, as well as other conclusions that emerged from the Group's discussion, convey the appearance of simplicity, but in fact were the product of much discussion and work to achieve a convergence of views. Put another way, the principles which serve to motivate and guide our work seemed quite logical and straightforward once unearthed; yet the process of unearthing these principles and arranging them in a useful way was often challenging in a group which brought to bear a variety of perspectives on the issues at hand.

5 Some open issues

In discussion, several interesting questions emerged.

Planning for change. A major concern flows from the sense of transition. We recognize that the emphasis and composition of library work is changing, and that to create value for users may involve de-emphasizing some areas and growing others. This raises the question of what one models: the library of today or what the library

needs to become. And this raises an urgent follow-on question: what are the appropriate services for new needs?

Reach and coordination. Our sense of urgency was corroborated by the fact that these questions have crystallized in other discussions and initiatives in our space:

- The NISO Blue Ribbon panel recommended that NISO needed an architecture or framework within which to align activities and priorities. It should not be merely reactive, but should try to shape its agenda within the context of a worked-out framework or architecture.
- OCLC/RLG/CNI/ALA recently held an invitational workshop of stakeholders in the digital library standards arena. One conclusion was that a framework which 'framed and named' the environment and its components would help in identifying gaps and priorities.
- JISC in the UK, DEST in Australia, and ADL in the US are collaborating on ELF, a framework for e-learning which is likely to be generalized to other domains.
- CIDOC, INDECS and FRBR are all advancing business entity models which overlap in scope.

But the multiplicity of initiatives directly or indirectly related to the Group's work suggests a need for liaison and coordination, if we are to avoid duplicative activity, inconsistencies, and ambiguous points of contact. It remains to be determined where and in what form this coordinating activity should be manifested.

Consensus and leadership. We believe that it is important that this type of work be done by those who have the most to contribute. This means that we need to develop a framework which draws on the creativity and vision of our community and beyond. It is important to achieve buy-in, but buy-in based on clear direction rather than lowest common denominator. We also believe that this work should spring from the real service imperatives of the library community. Managing the sometimes conflicting imperatives of, on the one hand, building a critical mass of stakeholder interest, and on the other, maintaining a sharp focus on clearly articulated library needs will remain a key issue as the Group proceeds with its discussions.

6 Recommendations moving forward

The Group has made considerable progress in the last few months, but at the same time, it has become clear that the current effort is inadequate to meet our goals. In particular, the Group feels that greater logistical support is needed for the Group's activities, and that dedicated attention must be allocated to this project in order to ensure steady, focused progress. Along these lines, the Group offers the following recommendations moving forward.

Two issues shape these recommendations: (i) this work is of central importance to the future of the research library, and (ii) it needs to receive serious attention.

The second of these points suggests that we need to develop a process which does not rely on resources solely sliced from the crowded agendas of busy people. Rather, we should have some dedicated effort which ensures coherence around goals and coordination of effective inputs. Furthermore, DLF should seek appropriate collaboration to move this forward.

Recommendation 1:

- The DLF should continue this work on the foundation that has been established.
- The DLF should secure dedicated effort to ensure coherence around a central vision and coordination of effort, drawing in expertise and results of other projects as appropriate.
- The DLF should explore collaboration with other bodies in the space to see if the cost of this effort could be shared. It may be that several bodies might be interested in providing some funding, securing a resource which has the authority to exercise the role outlined below.

Organization:

- The existing Group should continue as a sponsor group, subject to review of its composition. The role of this group is to approve and to guide.
- The Steering Committee should authorize the Director to secure the services of a DLF Fellow to coordinate the work moving forward. We recommend that the Director present funding strategies for this position, including (if need be) use of the Capital Fund. The Director would need to present a case for approval. The SF Group and the DLF Steering Committee would work with the Director to identify shared funding opportunities. This should proceed rapidly.
- This Advisory Group should be tasked with drawing up a person specification and with making recommendations.
- The role of the Fellow would be to:
 - Ensure consistency of the Framework.
 - Liaise with relevant initiatives to ensure appropriate participation/input.
 - Manage participation of groups.
 - Work closely with the Director to liaise with relevant DLF initiatives, such as Aquifer.
 - Liaise with ELF (the JISC/DEST e-learning framework outlined in Appendix D).

Structure:

- The work should initially be partitioned within 'business requirements'.
- In each, a small group should move along definitions of business functions (as defined within the Framework).
- The work of each group should be within a generally established framework, overseen and evolved by the Fellow supported by the Advisory Group.
- The work of each small group should be folded into the general framework with appropriate modification.

The result would be (i) a framework (ii) a method for creating and maintaining a framework (iii) a group of engaged people and (iv) an agenda of development work identified as gaps.

Recommendation 2:

Research Libraries are going through foundational changes. There is much discussion about futures, about flux, about value.

We recommend that the DLF commission work which develops some scenarios of how the future research library will deliver value in the research and learning lives of their users.

We recommend that this work be informed by participation from non-library expertise (business planning, systems architecture, supply-chain management/logistics). And that it include the perspectives of those to whom the University Librarian is accountable.

We further recommend that this work be pursued in association with the Mellon Foundation, who has expressed some interest in these topics, and with CNI, which has undertaken to look at this issue following recent discussions.

The aim is not to design the future library, rather to expand the range of thinking about alternative and desirable futures. The outcome would potentially be several scenarios, with accompanying discussion of direction and consequence.

7 Appendix A: The group's progress to date

The Group determined that a digital library abstract service framework should help to:

1. understand and communicate the 'business' of libraries,
2. support interoperability, portability and reuse of services and software to promote economies and efficiencies in development.

The first is particularly important if the value of libraries is to be understood; the second is particularly important if the value of libraries is to be realized.

More specifically, the Group determined that a digital library abstract service framework will:

- Make the contours of the digital library domain more explicit, to facilitate communication with other stakeholders.
What services can the learning management system, the search engine, the lab book, the RSS aggregator expect to consume from the library?
- Facilitate the design and development of complex digital library environments.
Without some partitioning of the problem space, it is difficult to design complex systems across different organizational, service, and technical boundaries.
- Guide the integration of services with user environments, whether those are personal environments or mediated by some other systems framework.
An explicit set of services clarifies communications and design decisions. Agreed interfaces support integration.
- Help align the supply and demand of services between domains (for example, the library aligns supply and demand of services across the publishing and user domains).
A broker role benefits from clear understanding of services offered and services required. Clean interfaces support the development of value added services.
- Support the development of services which may have unpredictable future users or uses.
- Help leverage legacy investment and guide future investment.
- Facilitate discussions between management and developers, providing a bridge between the 'business' of libraries and the 'systems' that support that business.

Our dominant model of library use is one in which library users approach a web-based user interface. Typically, a library will offer a variety of web-based interfaces, sometimes with an attempt to provide integrating portal or metasearch services.

At the same time, library users use other services, developed within other domains*.

* We understand a domain to be a social organization of shared practices and perspectives. Domains can also be thought of as 'knowledge interchange communities', or in other words, communities which attend the same conferences!

It was taken as an assumption that we (the group) operated within the library domain, but variably participate in other domains also. It was also recognized that these domains overlap in interest and participation. The group formulated a list of domains of primary importance:

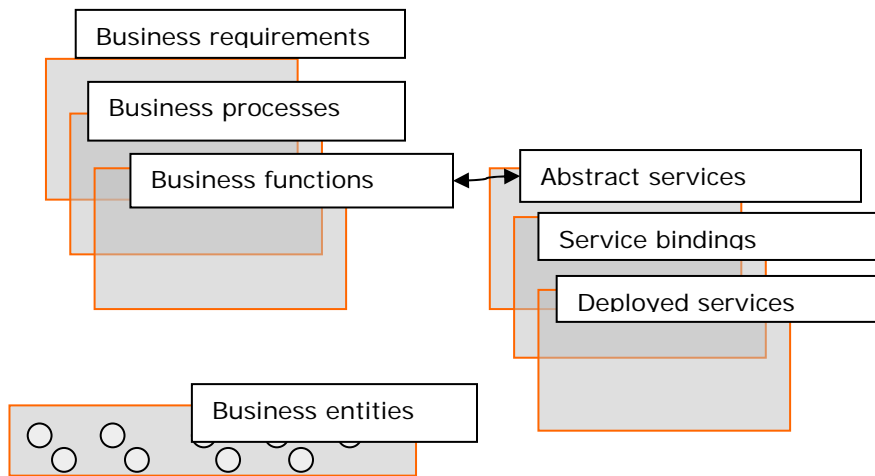
- E-learning (the management of learning in a network environment)
- E-research (this includes both emerging generic practices [see the discussions around grids, cyberinfrastructure and e-science for examples] and the growing work within particular subject domains)
- E-archives and e-records managers
- Publishing (considered loosely here, as the range of publishing activity grows wider)
- Enterprise systems (within campus environments)
- Personal users (in recognition of the growing importance of personal environments for consumption and generation of resources)
- Search engines and other open web services.

The importance of these domains is that within each there is a similar exploration of new network service environments which is creating new workflows, 'learnflows', and 'researchflows'. These flows are potentially the principal 'consumers' of library services in the future.

Thus, in line with the double emphasis noted above we can note parallel emphases here also: it is important that (i) these domains understand the services libraries have to offer (ii) those services are made available in ways that facilitate their reuse in multiple workflows and service environments.

After much discussion, and some exploration of services, we have decided to explore the two emphases we identified above as separate steps. This is so that we can understand the business requirements for service development in advance of thinking what services and articulations of them meet these business needs. As noted above, a major issue for us is 'business flux' in our changing environment. The Group sketched out a broad model of business requirements and service levels that will underpin the development of the abstract service framework (see Appendix B). The Group did some preliminary testing of this model in the context of two specific business requirements: discovery to delivery, and content management.

8 Appendix B: Model of business requirements and services



Understanding and communicating the business functions:

- A **Business Requirement** is an identifiable segment of an organization's overall mission.
Example: Discovery to delivery (D2D), collection management, ...
- A **Business Process** is an identifiable portion of a business requirement.
Example: Discovery, repository management, ...
- A **Business Function** is an identifiable portion of a business process
Example: Search

At this time we recognize that this division may be a little arbitrary: there are no hard and fast rules about demarcation. We expect consensus to emerge based on developing experience of working with the model. The aim is to develop a set of business functions. These are broad labels for library activities.

Sharing services and software:

- An **Abstract Service** is a conceptualization of business functionality as a discrete piece of networked functionality, consisting of a description of its functional scope and an abstract model of its behavior and data.
Example: Search
- A **Service Binding** is a specific instantiation of an Abstract Service. A Service Binding elaborates on an Abstract Service by providing all of the following which are applicable: 1) a specific data representation; 2) an application programming interface (API) specification; and 3) a Web service specification.
Example: WSDL description of an online search service implemented as a Web service.
- A **Deployed Service** is a Service Binding available at a specific location on the network.

Finally, we recognized that we needed to identify the 'business entities' of interest.

- A **Business Entity** is an entity of interest within the environment. Business entities are typically represented by metadata. Libraries are used to managing information entities; increasingly they will have to manage representations of other entities to create viable digital environments.
Examples: service, person, organization, metadata schema, terms & conditions, collection, policy, transaction, budget ...

We have also begun some work on a shared set of terms for operations, again to facilitate communication.

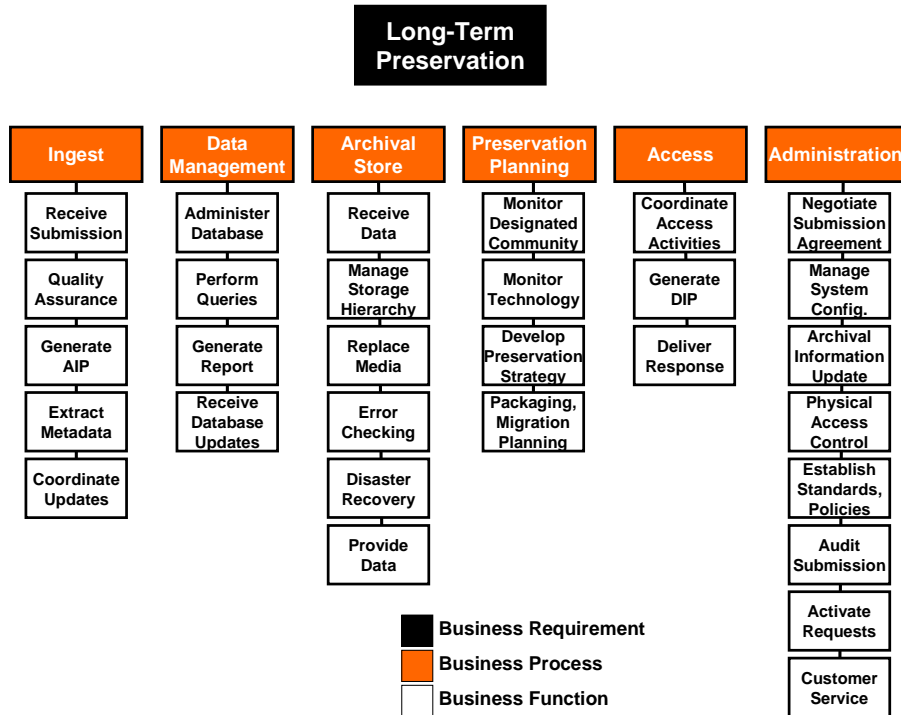
We anticipate that there is a many-to-many relationship between business function and abstract service. What connects them is an application.

A reference model may be developed at any level. A reference model is a particular articulation of entities to support a business goal. In practice, a reference model may be of most use at the business function level. For example, a reference model could show how a set of business functions supports the discovery to delivery requirement, or how a set of services is articulated within the collection management process.

This schematic needs to be refined and tested in actual examples. We have done some work in two business requirement areas: discovery to delivery and collection management. There are compelling points of contact between our work to date and any design work on Aquifer.

9 Appendix C: Business requirements, business processes, and business functions in practice: the OAIS reference model

To make the model described in Appendix B more concrete, below is an example of how the model's concepts might be applied to some existing work – the Open Archival Information System (OAIS) reference model. This is not intended to be a precise or authoritative application of the model to OAIS, but only an illustration of how the model relates to a real-world problem space.



In the example above, the business requirement is long-term preservation, which, according to OAIS, is an aggregation of six distinct business processes: ingest, data management, archival store, preservation planning, access, and administration. Each of these business processes is decomposed into multiple business functions. It is these business functions – or pieces of functionality – that serve as points of contact between the business and the services needed to support the business. For example, the ingest business process includes a business function called 'extract metadata'. It is easy to imagine that the functionality to 'extract metadata' could be mapped to an abstract service, defined as a piece of functionality that accepts as input a document in a particular digital format, and outputs a set of populated metadata elements. This abstract service could then be implemented as a service binding, which might include metadata schema, web service, and API specifications.

10 Appendix D: Related work

We surveyed a range of reference models, architectures and service frameworks, roughly grouped according to domain. We identified some consistency of motivation along the lines we have identified above, but no consistency of formulation. Most of these initiatives were attempts to shape discussions about services in particular domains or problem areas. Some were also intended to guide development work.

In the information environments domain, a number of architectures and models were available, including the Flexible Extensible Digital Object and Repository Architecture (FEDORA); the JISC Information Environment Architecture; the Making of America II Digital Library Service Model; the University of Arizona's Scholar's Portal work; the University of California's Common Framework, and the LANL aDORe work, which could be applied as an architectural foundation for cross-repository services.

The research and learning environments domain included the Content Object Repository Discovery and Registration/Resolution Architecture (CORDRA); the E-Learning Framework (ELF); the IMS Digital Repositories Specification; the Open Knowledge Initiative; and the University of California Berkeley's Scholar's Box. The preservation domain included the National Digital Information Infrastructure and Preservation Program (NDIIPP) Architecture and the Open Archival Information System (OAIS) reference model.

Some other resources noted by the Group were a White Paper on Enterprise Architecture from Denmark's Ministry of Science, Technology, and Innovation; the W3C's Architecture of the World Wide Web; and IBM's Open Grid Services Architecture. The Group also noted several data models of interest – in particular, the National Library of Australia's Digital Collections Manager Functional Specifications and Harvard University's Reference Model for Digital Library Objects.

These and other resources provided the Group with a useful survey of the current state of the art in architectures and reference models. Unfortunately, while we noticed considerable consistency of motivation, there was no immediately obvious model to adopt.

One initiative of particular interest is ELF (e-Learning Framework). This is an initiative jointly advanced by JISC in the UK, the Department of Education, Science and Technology in Australia, and some ADL interests in the US. Initially confined to e-learning, it is now seeking to expand its range of interest.

The focus of ELF appears to be on services, and reference models which combine a set of services in support of a particular business goal. ELF has not yet produced any reference models, although this is the subject of some funded work.

We have established lines of communication with ELF and its successor initiatives. We hope to be able to work together where it makes sense. ELF is still in early stages of development.

There are various other initiatives within the DLF community and beyond which are of relevance. None of them has quite the scope of this one, but several are relevant to particular parts of the picture.