1. Purpose, Goals, Objective, Merit, and Broader Impact

Our **purpose** is to ensure that people and institutions better manage information through digital libraries (DLs). Thus we address a fundamental human and social need, which is particularly urgent in the modern Information (and Knowledge) Age. Our **goal** is to significantly advance both the theory and state-of-the-art of DLs (and other advanced information systems) **-** thoroughly validating our approach using highly visible testbeds. Our research **objective** is to leverage our formal, theory-based approach to the problems of defining, understanding, modeling, building, personalizing, and evaluating DLs. We will construct models and tools based on that theory so organizations and individuals can easily create and maintain fully functional DLs, whose components can interoperate with corresponding components of related DLs.

This research should be **highly meritorious intellectually**. We bring together a team of senior researchers with expertise in information retrieval, human-computer interaction, scenario-based design, personalization, and componentized system development – and expect to make important contributions in each of those areas. Of crucial import, however, is that we will integrate our prior research and experience to achieve breakthrough advances in the field of DLs, regarding theory, methodology, systems, and evaluation. We will extend the 5S theory, which has identified five key dimensions or constructs underlying effective DLs: Streams, Structures, Spaces, Scenarios, and Societies. We will use that theory to describe and develop metamodels, models, and systems, which can be tailored to disciplines and/or groups, as well as personalized. We will disseminate our findings as well as provide toolkits as open source software, encouraging wide use. We will validate our work using testbeds, ensuring broad impact.

We will put powerful tools into the hands of digital librarians so they may easily plan and configure tailored systems, to support an extensible set of services, including publishing, discovery, searching, browsing, recommending, and access control – handling diverse types of collections, and varied genres and classes of digital objects. With these tools, end-users will for be able to design personal DLs.

Testbeds are crucial to validate scientific theories and will be thoroughly integrated into SciDL research and evaluation. We will focus on two application domains, which together should allow comprehensive validation **and increase the significance of SciDL's impact on scholarly communities.** One is education (through CITIDEL); the other is libraries (through DLA and OCKHAM). CITIDEL deals with content from publishers (e.g, ACM Digital Library), corporate research efforts (e.g., CiteSeer), volunteer initiatives (e.g., DBLP, based on the database and logic programming literature), CS departments (e.g., NCSTRL, mostly technical reports), educational initiatives (e.g., Computer Science Teaching Center), and universities (e.g., theses and dissertations). DLA is a unit of the Virginia Tech library that virtually publishes scholarly communication such as faculty-edited journals and rare and unique resources including image collections and finding aids from Special Collections. The OCKHAM initiative, calling for simplicity in the library world, emphasizes a three-part solution: lightweight protocols, component-based development, and open reference models. It provides a framework to research the deployment of the SciDL approach in libraries. Thus our choice of testbeds also will ensure that our research will have additional benefit to and impact on the fields of computing and library and information science, supporting transformations in how we learn and deal with information.

1.1. Impact on Libraries

Today many students and scholars cannot effectively manage their knowledge resources, have little training in library and information science, and lack understanding of organizing principles of scholarly communication, which is in upheaval as a result of a revolution in networking and electronic publishing. Library services have always existed for the benefit of such information-seeking communities, and in recent decades have evolved rapidly to afford effective use of new information systems. But this evolution has lacked the guidance of a science that formally models and tests theoretical understanding. Clearly, a science of DLs would have enormous beneficial impact on libraries as institutions, and on the communities they serve. SciDL will engage scientists and librarians in a collaborative process to model DL systems theoretically and then evaluate these models through testbeds. Through SciDL we will integrate across the full information cycle [18], first expanding and then testing the 5S model in a prototype and then in living library environments [22]. Building upon 5S, and toward a science of DL, our research should help to give DLs some of the soul that is currently lacking in the online environment.

1.2. Impact on Education

SciDL will facilitate learning by providing support for knowledge acquisition and dissemination. Toward these aims, it must ease communication among components in separate DLs – for two reasons: First, it allows an individual to have separate collections for different purposes and to easily make links among those collections. Second, it allows an individual's collection to connect with collections developed and maintained by others so each collection is richer for its association with a wider community. A well developed set of tools, based on solid scientific principles, will allow scholars to collect, organize, browse, and share digital resources as conveniently as they currently manage their other materials.

2. Approach

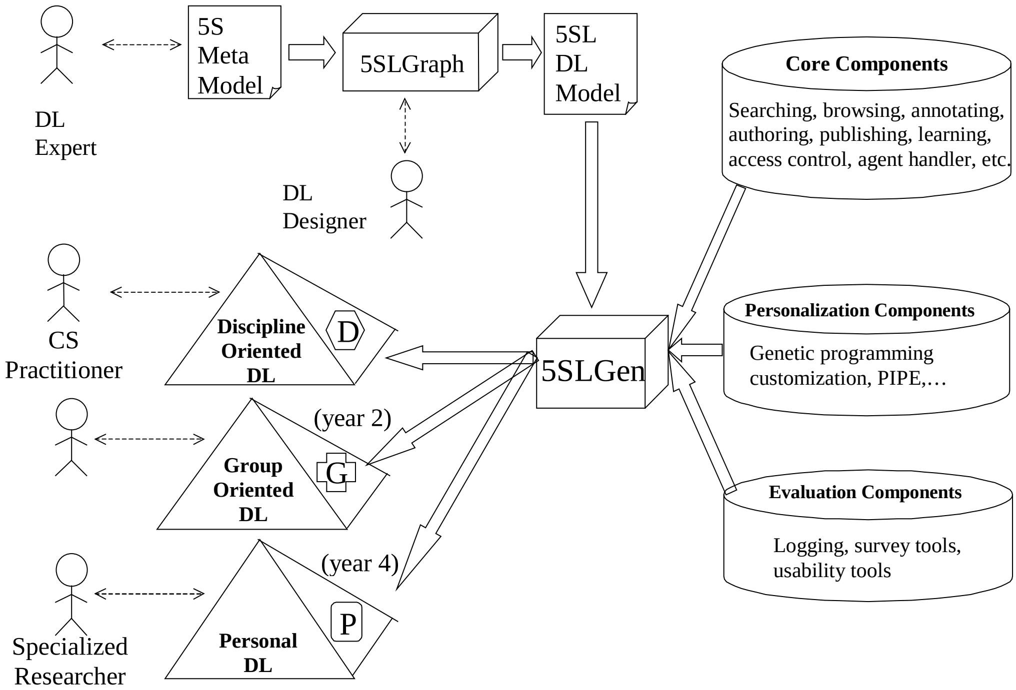

Figure 1. Overview of architecture for DL modeling, generation, and personalization.

2.1. Overview

As is shown in Figure 1, we adopt an approach shown to be highly effective in other areas of computing: develop powerful theories and (meta)models (i.e., 5S); use them to develop formal specifications (i.e., 5SL); generate tailored systems from those specifications (using 5SLGen); and integrate advanced methods of human-computer interaction (e.g., scenarios, personalization).

2.2. The 5S Framework for Digital Libraries

In this research, we propose a formal, theory-based approach to the problems of defining, understanding, modeling, building, personalizing, and evaluating DLs. We use mathematically based formal methods to develop our theoretical framework for DLs. Formal models and theories are crucial to specify and understand clearly and unambiguously the characteristics, structure, and behavior of complex information systems. It is not surprising that most of the disciplines related to DLs have underlying formal models that have served them well: databases [23-28], information retrieval [29-34], and hypermedia [35, 36]. A formal model abstracts the general characteristics and common features of a set of systems developed for similar problems, explains their structures and processes, and strengthens common practice. Furthermore, formal models for information systems can be used for the design of a real system, providing a precise specification of requirements against which the implementation can be compared for correctness. The use of formal methods does not *a priori* guarantee correctness, but can greatly increase understanding of a complex system by revealing inconsistencies, ambiguities and incompleteness that might otherwise go undetected [37]. The lack of formal models also can lead to diverging efforts, and thus may have contributed to making interoperability one of the most important DL problems.

2.2.1. The 5S Model

Recognizing difficulties in understanding, defining, describing, and modeling DLs, we introduced the 5S model, balancing rigor and usefulness [38]. Streams are sequences of arbitrary items used to describe static and dynamic content. Structures can be viewed as labeled directed graphs, which impose organization. Spaces are sets with operations on those sets that obey certain constraints. Scenarios consist of sequences of events or actions that modify states of a computation in order to accomplish a functional requirement. Societies are sets of entities and activities and the relationships between and among them. (N.B. Work on personalization is fundamental to a deeper understanding of Societies.) Together these abstractions provide a formal foundation to define, relate, and unify concepts -- among others, of digital objects, metadata, collections, and services -- required to formalize and elucidate "digital libraries".

2.2.2. The 5SL Language

In order to realize the potential of 5S to model and build DLs, we introduced 5SL, a domain-specific language for declarative specification and generation of DLs [39]. 5SL enables high-level specification of DLs in five complementary dimensions, including: the kinds of multimedia information the DL supports (Stream Model); how that information is structured and organized (Structural Model); different logical and presentational properties and operations of DL components (Spatial Model); the behavior of the DL (Scenario Model); and the different societies of actors and managers of services that act together to carry out the DL behavior (Societal Model). 5SL is an XML realization of the 5S model with specific considerations of interoperability and reuse built into its design. See details in Table 1 below.

Table 1. Theory-driven research and testbeds.

Theory			**Testbeds**	
Model	*Primitives*	*Objectives*	*Education*	*Library*
Streams	Text; video; audio; picture; software program	Describe properties of the DL content such as encoding and language for textual material or particular forms of multimedia data	Simulations; videoconference	Ejournals, ETDs, course materials, imagebase, EAD finding aids
Structures	Collection; catalog; hypertext; document; metadata; organization tools	Specify organizational of the DL content	Syllabi; lesson plans; hypermedia presentations	Classification schemes; ontologies; controlled vocabularies, name authorities
Spaces	User interface; index; ranking function; retrieval model	Define logical properties and presentational views of several components	Virtual classrooms; collaborative environments	Customized GUIs; personal search engines; virtual shelves; portals
Scenarios	Service; event; condition; action; state	Detail the behavior of the DL services	Authoring; annotating; simulating	Cataloguing; seeking; workflow; publishing; validating, research, instruction
Societies	Community; managers; actors; relationships; attributes	Define managers, responsible for running DL services; actors, that use those services; and relationships among them	Teachers; Learners	Researchers; authors; students, faculty, genealogists

2.2.3. 5SGraph Visual Modeling Tool

Modeling such a complex system using only an XML-based language requires a great deal of knowledge of the 5S theory and language syntax. The process can be cumbersome and error prone. When large and complex DLs are to be built, it is hard even for experts to manually write desired XML files without any assistance from a tool. Accordingly, we introduced 5SGraph, a visual modeling tool that helps designers to model a DL instance without knowing the theoretical foundations and the syntactical details of 5SL [40]. Furthermore, 5SGraph maintains semantic constraints specified by a 5S metamodel and enforces these constraints over an instance model to ensure semantic consistency and correctness. 5SGraph also enables component reuse, to reduce the time and efforts of designers.

2.2.4. 5SLGen Environment for Generating Digital Libraries

5SLGen is a DL generation environment, that combines theory, language, and tools in a coherent and cohesive way to allow automatic generation of tailored DLs. The 5SLGen environment and generation process is described in Figure 1. High-level DL conceptual abstractions and their properties are described in a *metamodel*, based on the 5S theory. A DL expert creates a metamodel for DLs and feeds the metamodel to the 5SGraph modeling tool. The modeling tool processes the metamodel, allowing the digital librarian (or the DL designer) to visualize the components of the metamodel. The visualization of the metamodel helps the designer understand the structure of a generic DL and reduces the learning time. The digital librarian interacts with the 5SGraph modeling tool to describe his own DL model, based on the specific requirements and needs of the societies to which the DL is targeted. The designer uses suitable parts of the metamodel to put together the final model of his own DL. The DL requirements acquired with the graphical tool are then formally captured using the 5SL domain-specific language.

The resulting 5SL models are then fed to the 5SLGen DL generator. The generator, armed with a powerful pool of DL components (from VT, ODU, and others, see section 2.3) generates a tailored, running version of the DL. This version of the DL can be further customized for particular individuals and communities using the personalization techniques proposed in section 2.5.

2.3. Componentized Architectures

2.3.1. Virginia Tech Components

We propose further development of componentized DL systems. There are two approaches we have been exploring, so we can carefully compare and contrast the types of systems commonly developed.

MARIAN: In the first approach, we build upon over a decade of work with the MARIAN DL system [41-47]. Through analysis of its object-oriented code base, we have assembled a component pool. Then, to build tailored DLs, we generate tailored MARIAN systems from specifications.

Open Digital Libraries: In our second approach, we draw heavily upon work with the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) [48]. In particular, we argue for Open Digital Libraries (ODL [49, 50]). This involves having well defined components that each carries out a key DL function, in whole or in part, and using extensions to OAI-PMH to connect components [52]. Such lightweight protocols ensure that different component implementations can be treated like interchangeable parts, and support inter-component interaction within a DL. In particular, DLs can be built by connecting small components that communicate through a family of lightweight protocols, using XML as the data interchange mechanism. Components adhering to these protocols were implemented and integrated into production and research DLs [49, 52, 53]. This work pre-dates current use of Web Services [54], which will be integrated with ODL as technology advances to ensure good performance.

2.3.2. ODU Components

In 5SL the DL expert designs the metamodel of DLs and the DL designer composes the model for a DL specific to the needs of a community. The DL designer must be aware of functional requirements - what services a community needs and what form of interaction these services should have with the users of the DL: publishers, searchers, and administrators. Customization includes both personalization, i.e., tailoring according to the needs of individuals, and selective functionality composition. Personalization includes both a one-time component done statically (usually at the instantiation of the DL) and a dynamic aspect that changes a profile as a user interacts with the system. In this section we describe, at a fairly high level, personalization objectives for selected services (components in 5SL terms).

Publishing: An early deliverable will be minimal-installation-cost publishing tools that allow us to test the concept quickly, using participants (project faculty and students) to create a very simple (i.e., it is extremely easy to publish into it) DL that can be harvested by any existing harvester such as Arc [55] or Kepler [56] using an agent handler. 'Minimal' refers to a system that can be installed on a personal computer without having specialized administrative privileges. The publishing tool will understand any XML schema for objects ranging from simple papers (PDF files) to objects that represent an entire course, including syllabi, lectures, tests, assignments, and projects. The latter might contain hundreds of files and references to outside web pages. The tool will understand the spatial XML schema and present the publishing user with an interface presumably geared to maximize ease of use for the specific objects published. The tool will interact with the repository component that actually stores the object in a database.

In the next phase we will create a version that can be installed on a server. Users will have the choice of 'Internet café' publishing while on travel or publishing on their personal machine. The publishing tool will include a mechanism to allow a publisher to build her own personalized bibliography to allow for easy construction of reference sections. The references will be stored as separate metadata that can be used by a reference linking service. Initially, we will develop workflow tools directly in the publishing service. They will be able to handle preferences of groups, to implement management of publication and review/annotation process of objects and be controlled by an access control system (see 2.3.2.3). We will then follow through with abstraction of workflow, and automation of generation of system components.

Discovery: We will adapt existing search tools to work with XML schema for user interface specifications and allow for customization within the tool for various forms of post-processing of the result set. The search engines will exploit the latest efforts in the arena of the semantic web to help users navigate result sets. These include refinement of the result set by author, date, and subject (classification). It will include a search of similar authors and similar subjects from either the entire collection or just the result set. Reference linking and citation services will be available from this navigation page, and flexible visualization will facilitate use by various groups [57].

Access Control: Almost all DLs have some form of access control, even if only to control who can publish into their collection. The ODU team has experience in using a distributed access control management scheme, Shibboleth [58], for its Archon (part of the NSF funded NSDL library) [59]. So, as we develop a system that can be personalized in a number of ways, we will allow for differential granularity of access, both in terms of objects and their complete structure, as well as users' roles within the organization (plurality of organizations if the DL to be constructed is distributed). This will be used to support workflow processes, as for example approving a document by the organization before supporting discovery. The access control component will be a core part of SciDL, i.e., it will be available for any service selected.

Agent Handler: Almost all modern DLs have provisions to deal with agents as well as with human users. One particularly important type of agent exists already for 5SL, the OAI-PMH handler. It will allow the DL to export the metadata of its objects. To this agent we will add the DP9 [60] system that will expose (under the discretion of the DL) the DL's metadata, supporting general web search engines such as Google. As part of this project we will develop a customizable DP9 that will flexibly support the DL designer.

Learning Services: We shall develop higher level services that support learning [61, 62], that is, publishing and annotation tools to 'grade' objects; assess their actual use from information about those who access them; and develop statistical reporting and analysis tools about effectiveness of DL infrastructure towards learning. The latter service will work from the data the logging service will produce.

2.4. Scenarios

Prior research has identified a set of key issues, but these have not been articulated or investigated with respect to DLs, or in particular with respect to DL interactions in education and libraries. For example, research in usability engineering has produced a set of concepts and techniques for involving users in developing requirements and designs for new systems, but in general this work has focused on transaction, document preparation, and communication systems, not on information design retrieval involving digital collections [63]. We will extend this line of investigation by developing usability engineering concepts and techniques in the education and library application domains for DLs.

2.4.1. Scenario Research

Research in scenario-based requirements analysis and design has identified issues in scenario generation, in the management of scenario descriptions, in analysis of scenarios, and in bridging from scenario descriptions to software designs and implementations [64]. For example, in scenario generation, one often sees systematic biases, such as the so-called representation bias [65]. We will continue to develop scenario-based system development concepts and methods through work in DL design. An initial direction is to adapt story-writing software (like Dramatica Pro; dramatica@screenplay.com) to support and integrate scenario development.

Research in computer-supported cooperative work has identified issues in the cost/benefit tradeoffs of collaborative interactions (e.g., the people in an organization who maintain shared calendars or send URLs to their colleagues are not the people who benefit from this work having been done; [66]), in the establishment and maintenance of awareness of the presence, actions, and activities of one's collaborators [67], and in the formation and development of trust and social capital [68-70]. Our prior work has investigated these issues in the context of Internet communities and (proximal) community networks. We will extend this line of investigation to learning communities.

Science and theory have been both rich and problematic in human-computer interaction. Sources of and contributions to science range from perceptual and motor analyses of human information processing, to theories of language and other cognitive/symbolic capacities, to formal specifications of interaction complexity, to the social psychology of groups, to sociology and workplace ethnography, to ethnomethodology [71, 72]. One approach to integrating this diverse science base is through design tools and reusable abstractions [73-75]. We will extend this line of analysis with respect to collaborative interactions of groups, learning communities, and other social organizations involving digital collections and tools for managing and using them.

Table 2. Example of a Scenario

Angela and Marty are working on a class project introductory physics. Angela is at Villanova University; Marty is at Virginia Tech. They are investigating whether smaller planets always move faster, and how the size and density of stars impacts possibilities for solar systems. They collaboratively examine and annotate images of solar systems, referring to audio annotations left by previous users, and then create a statistical analysis of all known solar systems. They then use the Alternate Reality Kit to simulate a series of solar system configurations, varying solar mass and density. Finally, they build visualizations of their results, and collaboratively write and edit a brief account of what they did and their conclusions. They make a specialized DL based on their research and presenting the collection they generate. The collection is intended to serve them later when they return to this topic, and also to allow others to benefit from the work they have done. Because their specialized DL is built to well established standards, they are able to export the metadata of their collection, enriching other related DLs. They also harvest metadata from sites they know are relevant, keeping their own collection well connected to emerging materials. They're in the midst of saying goodbye when they are contacted by a student from Old Dominion University who found their work while searching a related DL and has several follow-up ideas for all three of them to pursue.

2.4.2. Impact of Scenarios on Software Design

Carroll [75]argues for scenario-based design: (1) Scenarios are concrete (see Table 2) in the sense that they are experienced as low-fidelity simulacra of real activity, but (2) flexible in the sense that they are easily created, elaborated, and even discarded. (3) Scenarios keep design discussion focused on user activities, more specifically, (4) they keep design discussion focused on the level of task organization that actors experience in their tasks ("basic" tasks, in the sense of [76]). (5) This makes it easier for all stakeholders in a design, including end-users, to participate fully, and (6) creates a focal, use-oriented design representation that can be reused throughout the system development process for constructing prototypes and mock-ups, requirements analysis, use cases and software object models, user interface metaphors, design rationales, usability specifications, formative and summative evaluation test tasks, task oriented training, help and other documentation, etc. In [77, 78], Carroll argued that these attributes were just the ones required to address design problems, considered as a subtype of what [79] called ill-structured problem-solving.

2.5. Personalization

Personalization entails customizing information access, structure, and presentation to the DL end-user. Personalization is achieved in DLs if they afford complex, compelling, and user-adapted interactions. Studying how users interact with DLs, and understanding the frustrations they experience, provides ample motivation for personalization [80]. No deeply theory-based, or really sophisticated methodologies exist to build personalization systems; the SciDL project attempts to fill this gap. Central to our proposed approach is the PIPE methodology for personalization [81] (see Figure 2).

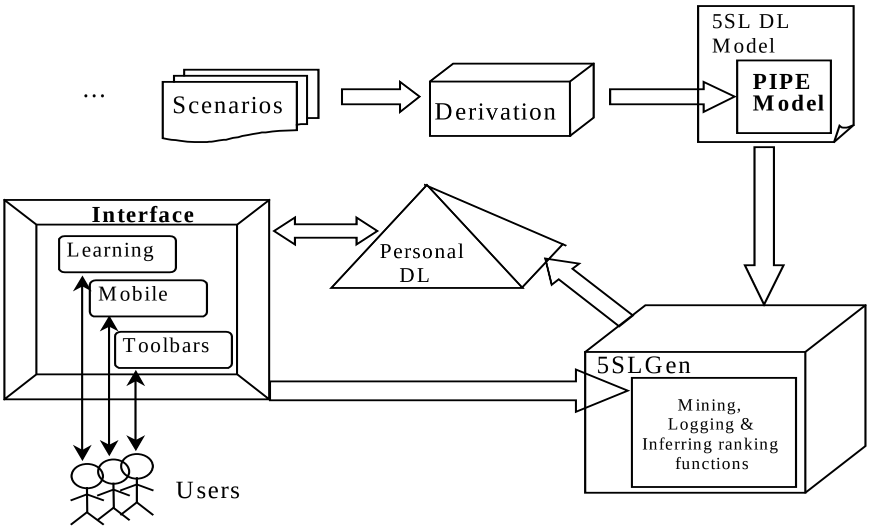

Figure 2. Personalization in DLs

2.5.1. PIPE: Personalization by Partial Evaluation

PIPE is an approach to personalizing information-seeking interactions by transforming programmatic representations. PIPE models personalization as a form of *partial evaluation* [82], an automatic technique for specializing programs given some (but not all) of their input. Representations in PIPE can be considered akin to scripts for summarizing information seeking [83] but are more amenable to transformation using commercial technology such as XSLT. Consider a PIPE model capturing navigation in the ACM Computing Classification system. For a given user, preference information about her interaction (e.g., her interests in graphics literature) is used to partially evaluate the PIPE model, resulting in a specialized navigation schema. This specialized schema can be rendered for browsing purposes or further specialized by another partial evaluation operation. A *sequence* of such partial evaluations thus corresponds to a personalized interaction. PIPE models can correspond to many navigational structures (e.g., hierarchies, web sites, 5S metamodels), allowing us to bring the full power of transformation to bear upon diverse information resources.

2.5.2. Deriving and Implementing PIPE Models

For PIPE to be useful in the SciDL setting, we must identify how PIPE models are derived and how the correspondence between the model and the DL is established (see Figure 2). One approach is to study interaction [80] from the viewpoint of the DL user's information seeking goals [84-86]. Scenario-based methods are ideal here in their role as empirical requirements analysis techniques. An alternate approach involves log mining to implicitly capture requirements. Ultimately these techniques are directed toward closing the gap between the goals of a DL designer and the task model of a DL user [87].

Since scenarios provide narrative accounts, techniques such as claims analysis and explanation-based generalization (EBG) are used to explain scenarios and provide a starting point for a personalization system. In particular, claims analysis can be used to extract individual interaction sequences from scenarios; EBG uses a domain theory (of task completion) to reason about the most important aspects of a scenario which have to be supported. For instance, explaining a scenario of a DL user interacting with a learning module can lead to the inference "Users prefer to browse online course material along topics featured in old exams". This inference leads to the creation of a PIPE model at the level of a topic hierarchy, cross-linked with exam metadata. Many such models will need to be created to support a range of interactions. Too many models will lead to a mushrooming of choices; too little would not be construed as "personalized enough." EBG and scenario-based methods allow us to formally and empirically study this tradeoff.

Once a PIPE model is derived, 5SL schema can be prototyped to support interactions implied by the model. This addresses data interchange formats and metadata specifications, as well as interfacing with the 5SLGen generator to create targeted DLs. As Figure 1 shows, such targeted DLs can be created at the level of a user, a subject domain, or a group of users. This synergy between PIPE and 5S, along with scenario capture (see Figure 2), provides a powerful vehicle for capturing, modeling, and realizing personalized interaction in a DL. Preliminary results are available in [88].

2.5.3. Personalizing Search Services

One of the important end-effectors for personalization involves specialized DL search services. To create such services in a DL context, we must support ranking functions at various levels of granularity, e.g., per-user, per-group, and per-subject. Furthermore, *consensus ranking* techniques [89] sometimes can be employed independently of the specific interaction. Results using such ranking techniques, while sharable among users, also pose problems when users demand more individual-tailored search results. For example, a novice and expert programmer, both looking for "java programming," would probably look for material written at different levels. Supporting this requires more effective user modeling and personalization of the search results – so-called personalized ranking [89].

To support more effective personalized information retrieval and discovery, we propose integrating into DLs the automatic generation of ranking functions using evolutionary search methods [90, 91]. The idea is to automatically combine different weighting heuristics to produce context-specific ranking functions for various search contexts (queries, users), based on feedback from users. This technique has been validated on various TREC and web corpora and the results are very promising. In SciDL, we will investigate the role of personalized ranking functions for various categories of DL users – learners, educators, and general users. It should be noted that personalization services should work along with the discovery components described in Section 2.3.2.

2.5.4. Integrated Personalization

We propose an integrated personalization framework that brings together the 5S DL model (Section 2.2), collections of components (Section 2.3), the scenario-based frameworks of requirements specification and software design (Section 2.4), the PIPE methodology for personalization (Section 2.5.1, 2.5.2), improved ranking functions for web access and information retrieval (Section 2.5.3), and more fundamentally, a practical understanding of DL usage contexts and the needs of emerging applications (based on the experience of all co-PIs).

A "lifecycle" of personalized DL system creation in SciDL can be summarized as follows (see Figure 2). An initial study of application domains, usage contexts, and testbeds leads to an initial 5S design for data formats, interchange specifications, and metadata modeling. End-user requirements and expectations are culled in the form of concrete scenarios and filed for further analysis. Tradeoffs and claims analysis are conducted to identify the most promising "starting points" for DL design and assessment. It is important to identify the most relevant IR/DL metrics at this stage so that the system can be engineered with these metrics in mind. Our aim is built-in evaluation, included from the creation of the system, not as an "afterthought." Operationalizing scenarios is an area we have investigated [92]. It yields prototype sequences of interactions that have to be supported – to cover a range of users and a variety of tasks. These sequences of interactions are then represented in a suitable 5S model and used along with PIPE to realize individual user interactions. At each step of this modeling, design decisions have to be made about the most suitable algorithmic approach to realize the IR/DL metrics. Issues such as combining multiple metrics, developing better ranking functions, and mining such information from user log data are relevant here. Simulations will yield performance data with adequate coverage and discrimination. An implemented DL system (using the 5SLGen DL generator) is then tested with real users. Results from this stage can lead to refining the starting set of scenarios, reconsidering the reification, and/or feeding back "live" data to improve the ranking functions and DL evaluation procedures.

2.5.5. Novel Interfaces for Personalized Interaction

We propose three main categories of interfaces to empirically evaluate the personalization solutions proposed here. To enable the PIPE approach to personalization, we will create interaction instruments [93] such as out-of-turn toolbars for customized browsing, and mechanisms to trigger logging [94], archiving, and mining of archival data. Such interaction instruments must integrate seamlessly with DL systems, both in terms of software architecture and from the end-user viewpoint.

Second, we plan to provide interfaces that are cognizant of device limitations such as screen size, memory, and input capabilities. For instance, if the user employs a PDA or a cell phone to access a DL, the 5S schema must be suitably enhanced to handle reformatting. Finally, targeted DL interfaces have to be realized for publishing, discovery, and learning services [61, 62]. Higher-level services are needed, i.e., publishing and annotation tools to: 'grade' objects, assess their actual use from information about those who access them, and develop statistical reporting and analysis tools about effectiveness of DL infrastructure towards learning.

3. Testbeds

SciDL collaborators will test and improve practical applications that enhance current and new library services. We will reduce barriers to information access for library users and library management, including reducing the number of staff hours required for labor intensive operations. New models and experimental components developed for SciDL will be able to function (and be tested and evaluated) in both controlled and live library environments.

3.1. DLA

Virginia Tech's DL and Archives (DLA, http://scholar.lib.vt.edu/), a service oriented and online resource wealthy department of University Libraries, will provide a fully functioning library environment in which to test our research for adaptability, functionality, and interoperability; to implement appropriate components; and to evaluate new services. It allows users to work in a live environment while we research application requirements, models, component services, and scenarios for communities, as well as personalization for individuals. SciDL also gives us the opportunity to evolve DLA to a higher level of DL (e.g., more compliant with OAI-PMH) – adapting a component model (e.g., [49]) while broadening scholarly communications functionality including editorial functions, workflow, and migrations from private / limited to full public access.

To achieve breakthrough impact, SciDL must develop software that can be deployed by non-computer scientists in a fully operational DL. DLA will adopt SciDL software for long-term use, replacing overly specific applications with flexible alternatives. For example, DLA is the home of the NEH-funded South Atlantic Humanities Center (SAHC) DL; as SAHC becomes established, appropriate SciDL applications will be tested, refined, and employed. This is an opportunity to deploy a model DL for regional humanities studies, built from the ground up with SciDL technology.

Since DLA contains information resources with varying levels of public availability, it will provide an excellent environment in which to test the distributed approach where library users (both patrons and authors) define and maintain their profiles. DLA provides the opportunity to test long term sequential users such as graduate students moving through masters to doctoral programs and authoring electronic theses and dissertations, a requirement at Virginia Tech since 1997. With log files back to 1996, it will also be valuable to compare and contrast patterns of use from pre- through post SciDL.

3.2. OCKHAM

OCKHAM (Open Component-based Knowledge Hypermedia Applications Management, http://ockham.library.emory.edu/), based at Emory University, broadly aims at improving the interoperability and affordability of DL systems. This initiative is affiliated with the DL Federation and also includes librarians, researchers, and systems professionals representing California DL, OCLC, Notre Dame, University of Arizona, University of Windsor, and Virginia Tech. OCKHAM proposes integration of lightweight protocols, collections of open components, and reference models. Reference models provide an effective mechanism for coordinating group activities in collaborative software development. Despite the extensive accumulated history of practice in libraries, there never have been systematic attempts to document the activities and services that libraries provide at the highest level of a reference model, especially as such models relate to new DL functions.

The OCKHAM subgroup of the SciDL consortium will develop three high level abstractions of academic library services, and the connecting points between these reference models. These reference models will be tentative initially, and subject to revision over the extended period of the SCIDL project, reflecting feedback that we will receive from others over time. The *Public Services Reference Model* will articulate the services that academic libraries provide to their user communities. *The Internal Services Reference Model* will describe the services that academic libraries provide to units internal to the library organization. Finally, the *Integrated Services Reference Model* will portray the relationships of library services (both public and internal) to other educational enterprise activities, such as courseware systems. Reference models documents will highlight existing protocols for intercommunication where they exist, and gaps where new protocols are needed. Reference models developed will be used to integrate DL components developed at other SciDL sites with existing library systems through interconnection components. The aim of all these efforts is to create an interoperable and clearly articulated software component infrastructure for library operations.

3.3. ODU

ODU will engage in a mixture of phased experimental prototyping and analysis. With totally new concepts, like 5S, it is difficult to predict all the issues and problems to be faced. Instead of concentrating on a full design and implementation of a complete system we propose to develop systems in phases, with continuous feedback from the community. For this purpose, we will develop a testbed at ODU with minimal functionality by the end of 18 months. To start with, the testbed would consist of an instantiated version of a DL that is based on 5S principles, driven by specifications of structures, streams, and spaces.

At every stage of our development we will receive feedback from users and will refine our design. In phase one we will deploy publication tools – archivelets – that can be personalized to handle simple digital objects typical for the publication needs for faculty. They will be tested with selected faculty in two departments at ODU. In the next phase, at the end of 24 months, we plan to extend the testbed to support configurable complex digital objects and open the DL to the student community as well. Typical objects specified in the testbed will be course materials, student portfolios, technical papers, personal bibliographies, and the like. This will provide the opportunity to test the use of our learning services (through assessment and evaluation) by these communities. At the end of 36 months, the testbed will support workflow processing, access control, and agent handling. At this stage, we will encourage the use of the DL testbed in the teaching environment. In the last two years, the testbed at ODU will expand to integrate the functionality being developed at other sites. In parallel, the library at VT will deploy ODU-built components and will provide feedback to ODU. The penultimate phase is to abstract all these specification into an advanced metamodel and to allow for automated DL generation. This will be deployed in the VT testbed and evaluated there. A fully integrated system will be deployed for the ODU community: faculty, students, and administrators. In the last year the integrated system will be deployed at the national level, including in other efforts of NSF such as NSDL.

4. Evaluation

4.1. Evaluating the Impact on Education

The tools produced through this project have the potential to enhance education by giving students and faculty the ability to create and share collections. We will assess how relevant the system is to the potential users, how many know about it, and how many actually incorporate it into their way of working. Evaluation will take several forms. Periodically during the term of the project, the goals and accomplishments of the project will be presented at public forums for user feedback. Goals will be revised as needed to reflect user interests and needs. Users will assess both the aims of the work and the degree to which the tools that are produced are easy to use and address their needs.

In addition to the regular feedback from local users at our four universities, through the CITIDEL project we benefit from widespread use of this DL educational testbed. Furthermore, we have experience with developing a standard for logging in DLs, based on the 5S theory, which will inform data collection in the current project [94].

4.1.1. Computing and Information Technology Interactive Digital Educational Library

CITIDEL[95] is a project funded through the end of 2003 within the NSF NSDL program that provides access to content from publishers (e. g., ACM digital library), corporate research efforts (e.g., NEC's CiteSeer), volunteer initiatives (e.g., DBLP, originally a bibliography on databases and logic programming), CS departments (e.g., NCSTRL, mostly technical reports), educational initiatives (e.g., Computer Science Teaching Center), and universities (e.g., courseware as well as theses and dissertations from the NDLTD worldwide consortium). A number of important resources currently exist in CITIDEL. They range from the extensive collections of computing literature in the DLs of ACM and IEEE-CS to lists of interesting web pages gathered and maintained by individuals. NSF and other funding organizations have supported creation of a wide array of resources, many of which would be of great value to teachers and learners if they were more widely known. CITIDEL serves as a portal (front-end) to all educational resources related to computing and information technology, aiming toward at least one million resources. Building upon work in the Open Archives Initiative [96], we have harvested metadata from all applicable repositories and provided integrated access and linking across all related collections. Considering the 5S framework, we apply the Open Digital Library and MARIAN digital library software developed at Virginia Tech, as well as ResearchIndex/CiteSeer and niche search engine technology from Penn State (and NEC), to develop tailored services for the various parts of our broad user community. CITIDEL involves diverse groups so that DL services and content can be tailored to aid the education efforts of the computing and information technology field. We will extend the CITIDEL throughout the course of this project to incorporate the newly developed DL tools for the purpose of enhancing education at all levels. The tools that will be produced in the new project will create new DLs that can be connected to CITIDEL if their creator wishes, thus enhancing the CITIDEL view of the world of resources in the domain of computing education (and NSDL), and providing greatly expanded exposure to the DL created by an individual or a group. Through conference presentations and focus groups, we will explore the response of intended users to the new functionality. Through logs, we will monitor user visits to the site.

4.1.2. Evaluation with Logs

Log analysis can identify how DL patrons use DL systems and services and how systems behave while supporting user information seeking activities [84-86, 97-101]. Log recording and analysis opens opportunities to improvements and enhanced services. We defined an XML based DL log format [94], using 5S as a guide for organizing the log structure and defining the semantics of user services and the DL components whose behavior is logged. Our initial aim was to support CITIDEL and other parts of the National Science Digital Library. This proposed standard led to a logging toolkit that can be plugged into any DL system (e.g., CITIDEL, MARIAN), thereby facilitating interoperability, reusability, and logging completeness. Part of the current project seeks to incorporate logging capability into all the components of the DL so all aspects of system behavior can be made visible for analysis and evaluation. A new, enhanced log separate component will collect information generated by other components; other routines will provide analysis and reporting capabilities configurable to the needs of each DL system.

4.2. Evaluating the Impact on Libraries

As SciDL sites generate components, a library-oriented selection of these will be deployed at DLA, Emory University, and optionally at other OCKHAM sites as testbeds for the three reference models. Generally, each testbed will address a real need, while demonstrating the utility of the SciDL framework. Integration activities should primarily take the form of programming small interconnection components to assimilate components from the various SciDL efforts. All such interconnection components will be released as open source tools. DLA will upgrade its current library services and resources with the results of SciDL throughout the life of the grant and will sustain these higher level DL resources and services after the grant is completed.

Library testbeds will be evaluated by a combination of usability studies, software trials, focus groups, and user surveys. The general aim of the evaluation process will be feedback for the SciDL consortium regarding the utility of the emerging science of DLs, as well as portions that need improvement. Evaluation reports for all testbeds will be prepared and disseminated via public presentations by the end of the project, as well as through appropriate publications and web sites.

DLA, as a testbed for all aspects of 5S, will collaborate: a) on publishing and searching with ODU (comparing and contrasting current and newly developed tools for publishing and federated searching), b) on personalization with Villanova and VT, and c) on reference modeling with Emory. DLA will evaluate the effectiveness of newly implemented components throughout the life of the grant and beyond. We will extend our current practice of online surveys of users—both authors/editors and researchers (i.e., library patrons). We will collect and analyze logs, comparing with DLA logs covering 1996 to the present.

5. Work Plan

This project follows the workplan shown in Figure 3. We break the entire project into its key parts: reference models, 5S/5SL models and tools, componentized architecture, personalization, scenarios, testbeds, and evaluation. Most of these parts can be implemented in parallel. The first two years of the project focus on the development of various reference models and testbeds for educational and library users. We then analyze these reference models using a scenario-based approach for various tasks and devise the PIPE models and 5S meta-models for them. In the same period, we also start the component building process to build the foundation for the component-based architecture. At the end of Year 2, we should have a relatively complete DL building system ready for deployment and testing in the subsequent years. In Year 3, we concentrate on the development of the personalization aspect of the 5S framework in order to generate personal DLs. We also revise the reference models and various 5S meta-models based on initial evaluations using the education and library testbeds developed in Year 1 and Year 2. By the end of Year 3, we should have all key components ready. In Year 4, we integrate the personalization component with the rest of the 5S framework. We will produce various DLs for different types of users and perform large-scale evaluations of these personal DLs using the testbeds we designed earlier. The evaluation of DLs for group users and community users will also be under way at the same time. In Year 5, we continue the evaluation process for education DL, library DL, and personal DL. We will also evaluate various 5S models and make necessary modifications on these models. Finally, we will develop personalization interfaces to support various mobile devices.

6. Related Work

An extensive DL literature has developed. Although space will not permit a detailed discussion of this literature, the co-PIs on this proposal have acknowledged the many contributions in a number of review articles and books [63, 64, 71, 72, 77, 102-105]. The most directly relevant works are summarized below.

5S: Theoretical, formal approaches are rarely reported in the DL literature. Wang's "hybrid approach" [106] defines a DL as a combination of a special purpose database and a hypermedia-based user interface.

		Year 1				Year 2				Year 3				Year 4				Year 5			
Reference Models	Partner	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	54	57	60
Build Library RM	E																				
Build Education RM	VT																				
Revise Reference Models	VT, E																				
5S/5SL																					
Models																					
Build Models	VT,O																				
Evaluate Models	VT,O																				
Tools																					
Build 5SLGen	VT,O																				
Build 5SLGraph	VT,O																				
Architecture																					
Build Components	VT,O,VU																				
Deploy components	VT,O,VU																				
Personalization																					
Derive PIPE models	VT,O,VU																				
Personalize Search Services	VT,O,VU																				
Integrate with 5SLGen	VT,O,VU																				
Develop Interfaces	VT,O,VU																				
Scenarios																					
Usibility Engineering	VT,E																				
Requirement Analysis	VT,E																				
Testbeds (Edu., Lib.)																					
Build Testbeds	VT,E,VU																				
Deploy Testbeds	VT,E,VU																				
Evaluation																					
Library	E																				
Education	VT																				
Personal DL	VT,O,VU																				

Legend used: **VT: Virgnia Tech; E: Emory; O: Old Dominion; VU: Villanova Univ.**

On-going work

Concentrated effort

Figure 3. Project Timeline

It builds upon this combination to formalize DLs in terms of the formal language Z [107]. Kalinichenko et al. [108] presented a canonical model for information systems and a compositional approach that provides a partial solution for interoperability in DLs. Castelli et al. [109] formalized the concepts of documents, based on the notions of views and versions, metadata formats and specifications, and have proposed a first-order logic based query language. Unfortunately, none of these is a comprehensive formalization.

Formal models precisely and unambiguously define the semantics of specific abstractions of a field. In the case of computer science, this allows exploitation and development of declarative approaches in design and development. Recent database research has been investigating declarative approaches and representations for specific kinds of information systems, mainly web and e-commerce sites. Strudel [110], Tiramisu[111], and Active Views [112], are examples of systems with this data-centric perspective of a web site. They all aim to separate Web site structure, data management, and page presentation.

The hypertext/hypermedia community has developed rich abstraction models and decompositions for systems, e.g., OOHDM [113], Web2000 [114], and Autoweb [115]. The WebML modeling language and its supporting tool, Torii [91], provide powerful abstractions to describe and generate the hypertext and navigation structure of Web sites. An even closer approach is described in [116].

ODU's Digital Library Definition Language (DLDL) focuses on describing external behavior of DLs to support interoperability in terms of federated searching. A similar approach is defined in [117]. Much attention has been paid to DL architectures and systems, e.g., monolithic systems (e.g., Greenstone [118], MARIAN [44, 47], componentized architectures (e.g, ODL [52, 53], OpenDlib [119], agent-based architectures (e.g., UMDL [120]), and layered architectures (e.g., Alexandria [121]). Similarly, though visualization in DLs supports search results (e.g., 2D-Envision [57, 122], 3D-SPIRE [123], hierarchical [124]), interaction with services [125], exploration across heterogeneous sources [126], query synthesis [127], and customization of interfaces for particular societies (e.g., for children, teachers [128]), we are unaware of any work like our domain-specific graphical environment for DL design and modeling.

Scenarios: Through the 1980s, scenarios began a march forward through functional specification documents. They appeared in the specification body, directly supporting definition of system components and functions. Then they appeared at the front of the document, conveying succinctly the design vision of the system: The traditional functional specification became the appendix. By the mid-1980s, scenarios were widely used as a working design representation [129, 130]. In the 1990s, scenarios supported design of computers and software [64]: Product planners presented day-in-the-life scenarios to their managers as design proposals [131, 132]. Requirements engineers gathered workplace scenarios through observation and interviews, and analyzed scenarios as primary data [133-138]. Scenarios helped specify functionality [139, 140]. Scenarios were employed as shared representations in design meetings [141-143]. System software was designed scenario by scenario [144-148]. Typical interaction scenarios were used as rubrics for system documentation, help, and user training [104, 140, 149], as well as for designing evaluation tasks [150, 151]. Scenarios were employed to codify and convey design knowledge [152, 153].

Scenarios are now pervasive in the design of computer systems / applications. Major design methods for human-computer interaction are scenario based [63, 154-156]. Scenarios support integrating human-computer interaction design with software engineering and strategic planning [157]. Scenarios support analysis of context of use in keeping with the ISO 9241-11 international usability standard [158]. A survey of industrial software development groups reported that *all* groups employed scenarios [159].

Personalization: Personalization technologies have become ubiquitous in all aspects of information delivery, access, and presentation. The problem faced by personalization research is symptomatic of the viewpoint expressed in this proposal, namely the lack of a unifying model or methodology for creating and fielding personalization software. As Riecken [160] points out, there are "personal views of personalization". Personalization of search products and DLs has been undergoing very active research recently. An ontology-based approach is proposed in [161]. A similar idea is discussed and implemented in the Outride system [89]. Besides the above standalone client applications, existing DL systems also try to augment existing DLs with some personalizable/customizable features to allow them to search and use DLs in a more user-friendly way such as customizable interface and data source collections [108, 162], individual-tailored search results, and information tracking [163, 164].

ODU's related work on Kepler (an NSF project)[56], Comopt (a US Navy project)[165], and Archon (an NSF project)[59] facilitates developing tools and processes that can be tailored at the component level. The customization aspects in these projects are at different levels. One example is the flexibility to support different complex digital objects for different communities. Another example is to have customizable search result presentations for end users. The ODU team also has the experience to customize workflow processes that are typically present in communities for publishing digital objects.

Villanova's Web Host Access Tools (WHAT) project deals with privacy and user centered web search. Our project will extend it to allow user profiles reuse as users enter new DL sites.

Education DLs: The Distributed Expertise project [17] established the extent to which course design and implementation is influenced by what instructors feel comfortable with. Finding suitable resources to support their educational activities is very much of interest, for example, in core computing courses where faculty often are assigned to teach outside their area of expertise.

The CITIDEL project, part of the NSDL, addresses that need for established resources by providing access to a large number of resources in a single location. CITIDEL provides a portal to a number of collections through a single interface. It allows searching and browsing many collections, keeps up to date with changes in those collections, and augments the discovery features with tools that support organizing and using what has been found. For the CITIDEL site, and all of NSDL, to reach their full potential requires that the community be empowered to create DLs of personal collections that easily interact.

7. Prior NSF Funding

Fox (VT). Computing and Information Technology Interactive Digital Educational Library (CITIDEL) NSF DUE (NSDL) 0121679, $800,000 over 2 years. A consortium led by Virginia Tech, with Hofstra, College of New Jersey, Penn. State, and Villanova, is building CITIDEL to serve the computing field. Content covers computer engineering, computer science, information science, information systems, information technology, software engineering, and all other related fields.

Carroll (VT). The School as a Knowing Organization — Knowledge Management as a Strategy for Continuous Teacher Development. NSF-REC-0106552, $710,223 over 3 years. This project applies knowledge management (KM) concepts and techniques, and information technology, to understand and enhance KM in schools. An organization's knowledge is inaccessible when and where it is needed, since it often is locally produced, haphazardly disseminated, and ineffectively indexed. Our TeacherBridge (Basic Resources for Integrated Distributed Group Environments) supports online resource sharing and other collaboration among high school science teachers in two Appalachian counties [166, 167].

Ramakrishnan (VT, recipient of NSF CAREER grant EIA-9984317). SGER: Personalization by Partial Evaluation. NSF IIS 0136182, $50,000. 2 years. Building on research in the areas of problem solving environments [168], recommender systems [169, 170], data mining [171, 172] and personalization [81, 93], we proposed the PIPE methodology of personalization, which equates personalization to the partial evaluation of a user's interaction with an information system. PIPE also connects personalization in web sites to mixed-initiative interaction functionality in voice-based architectures such as VoiceXML [168].

McMillan (VT). A Digital Library Network for Engineering and Technology (DLNET). NSF DUE (NSDL) 0085849, $605,573, 2 years. This project developed a Digital Library Network for Engineering and Technology, to facilitate the lifelong learning of engineering faculty, practicing engineers, and technical professionals. It developed a content hosting platform, standardized templates for posting new content, a process for electronic review and validation of new materials, and a content portal – regarding education and research materials published by universities and professional associations (e.g., IEEE).

Maly and Zubair (ODU). Use of digital libraries for Undergraduate Learning (UDLF). NSF IIS-9816026, $80,355, 1 year. ODU developed a model of digital objects useful in a higher education environment: entire courses, each component having suitable metadata to be used in a tailored search interface. We completed a prototype tool-set (publishing and searching) that used an XML specification of complex objects as a common structure and applied the methodology to implement the UPS prototype library of some 200,000 objects – a key step in the unfolding of the Open Archives Initiative [173].

Cassel (Villanova Univ.). MRI: Web Host Access Tools. NSF CISE/EIA 0079770, $136, 551 over 3 years. This project addresses important questions in artificial intelligence, information gathering, human-computer interaction and networking. The problem that joins these topics is assisting a user retrieving and using information obtained from WWW. Cassel is also a co-PI with Fox on CITIDEL.

8. Management Plan

As is shown in Figure 4, the management of the project involves three components.

1. The Executive Office will be at VT, made up of the PI (Fox) and a half-time Project Manager (to be hired from the rich labor pool in Blacksburg, a college town with many talented individuals looking for such opportunities). The Project Manager will: serve as a central point of contact for all team members, organize team meetings, maintain schedules and a calendar of presentations and other dissemination activities, manage a DL of project documents (built using SciDL components), and coordinate the flow of information among team members. Please also see attached letters of support from Mark McNamee (University Provost and Vice President of Academic Affairs, Virginia Tech), Thomas Isenhour (Provost and Vice President of Academic Affairs, ODU), Joseph Lucia (University Libraries and Director of Falvey Memorial Library, Villanova), John Johannes (Vice President for Academic Affairs, Villanova), and Carol Weiss (Director, Villanova Institute for Teaching and Learning).
2. The Advisory Board, meeting in person in conjunction with the leading DL conference (Joint Conference on Digital Libraries) and communicating electronically in the interim, will review and guide project efforts. Members of the Board are among the most distinguished researchers interested in DLs and related fields. (See names in Figure 4 of those who have confirmed their support of this project by way of agreeing to serve.) They run the leading DL research teams in universities, or represent key related institutions (the electronic and DL efforts of the first professional society in computing, one of the largest DLs in a national laboratory).
3. The Project Team will be led by 9 co-PIs, supervising a larger number of graduate students and staff. In addition to the work performed at VT, important contributions will be made by teams working at Emory University, Old Dominion University, and Villanova University.

The following paragraphs describe the work that each of the co-PIs will manage in the project.

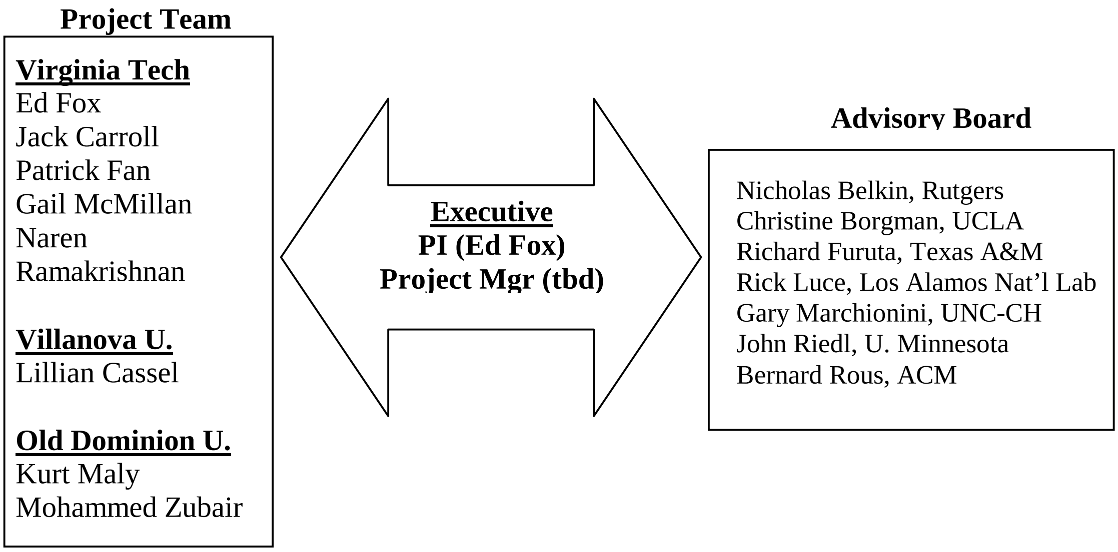

Figure 4. Management Plan

<u>Fox</u>: This research program will be directed by Professor Edward Fox, who has served as PI or co-PI on over 80 research grants. He led one of the first NSF-funded DL projects in 1991. Fox has collaborated with all of the co-PIs on other projects. He and McMillan co-founded the Electronic Thesis and Dissertation (ETD) project at VT, which has since grown to national and international dimensions. He has directed the Digital Library Research Laboratory at VT since 1998, and is Director of CITIDEL, the Computing and Information Technology Interactive Digital Educational Library, part of the National Science Digital Laboratory (NSDL). He is Chairman of the Policy Committee of NSDL. Fox became interested in this area 35 years ago, working with JCR Licklider (whose challenge to develop an integrative theory for libraries of the future in [174] led to 5S) and then with Gerard Salton [175].

Carroll will investigate collaborative responsibilities and requirements for groups, learning communities, and other organizations. His expertise covers usability engineering, scenario-based requirements analysis and design, computer-supported collaborative work, and human-computer interaction science / theory. He will lead efforts related to the Scenario part of 5S, and will play a key role in the Societies related efforts.

Fan will help with design and implementation of DL service scenarios, building upon five years of research on novel technologies for effective information retrieval and discovery, question answering, and summarization of search results. He will develop personalization techniques based on Genetic Programming that can be plugged into a DL, suited to an individual's preference.

McMillan brings over ten years experience with DL resources and services that will provide the connection between theory and practice in a fully functioning library environment. As a department head, she is a member of the University Libraries' Library Administrative Council. She will facilitate the use of her department of the University Libraries, the Digital Library and Archives (DLA), for testing and evaluation of theoretical models, as well as systems and components developed.

Ramakrishnan has worked with Fox and Carroll independently on various aspects of personalization. This research brings a fruitful meeting point for consolidating ideas towards the common goal of DL personalization. The PIPE software framework is maturing and SciDL will provide valuable testbeds for studying realistic personalization contexts. Ramakrishnan also collaborates with the University of Minnesota's John Riedl and Joe Konstan, and this work brings pertinent background in recommender system design and evaluation – critical facets for DL personalization system assessment. This has been a thorny issue in the recent past but declarative frameworks such as 5S and the use of transformation-driven personalization (i.e., PIPE) provide formal methodologies for addressing this problem.

Cassel at Villanova Univ. will focus on three areas – Personalization, Evaluation, and Educational Impact. The Personalization work will lead to a distributed profile facility that will allow individual users to retain the profile information that gives them personalized access to any DL on the first visit. Cassel will be responsible for coordinating the evaluation efforts of all the sites, generating clear checkpoints, and assessing progress on a regular basis. Evaluation results for the overall project will be reported at approximately the midpoint and the end of the project. Experience in analysis of user logs on the CITIDEL project will be utilized to make use of logging in all components of the new systems. Components in the analysis of Educational Impact will include establishment of the needs and interests of teachers and learners with regard to DLs, integration of components into the CITIDEL project, and querying user responses to these features. Cassel spent her 2002/2003 sabbatical largely at an office in Virginia Tech's Digital Library Research Laboratory.

Maly and Zubair at Old Dominion Univ. will contribute to the development of the 5SL schemas and develop publication tools that can be personalized for use by faculty. They will deploy a minimum functionality testbed for educational use by selected faculty within 18 months. This testbed will help develop the workflow process model for further elaboration of the tools. In later stages, the testbed will be enhanced with additional components and access will be expanded to include students. This iterative cycle will provide valuable input to other members of the team to guide component development. The ODU group will coordinate closely with the DLA activities at VT. The ODU team has worked with Virginia Tech since 1993 regarding DL support of computing technical reports.

Halbert at Emory Univ. will be responsible for developing three high level abstract models of academic library services – the Public Services Model, the Internal Services Model and the Integrated Services Model. He will bring the resources of OCKHAM, a consortium of professionals involved in the development of DL systems, into this effort. He will establish a library testbed to deploy DL software developed by other members of the SciDL collaboration and evaluate its use by clients in a real life library setting. Feedback from the testbeds will allow further development and refinement. Halbert has worked with Virginia Tech since 2001 on AmericanSouth.org and OCKHAM.