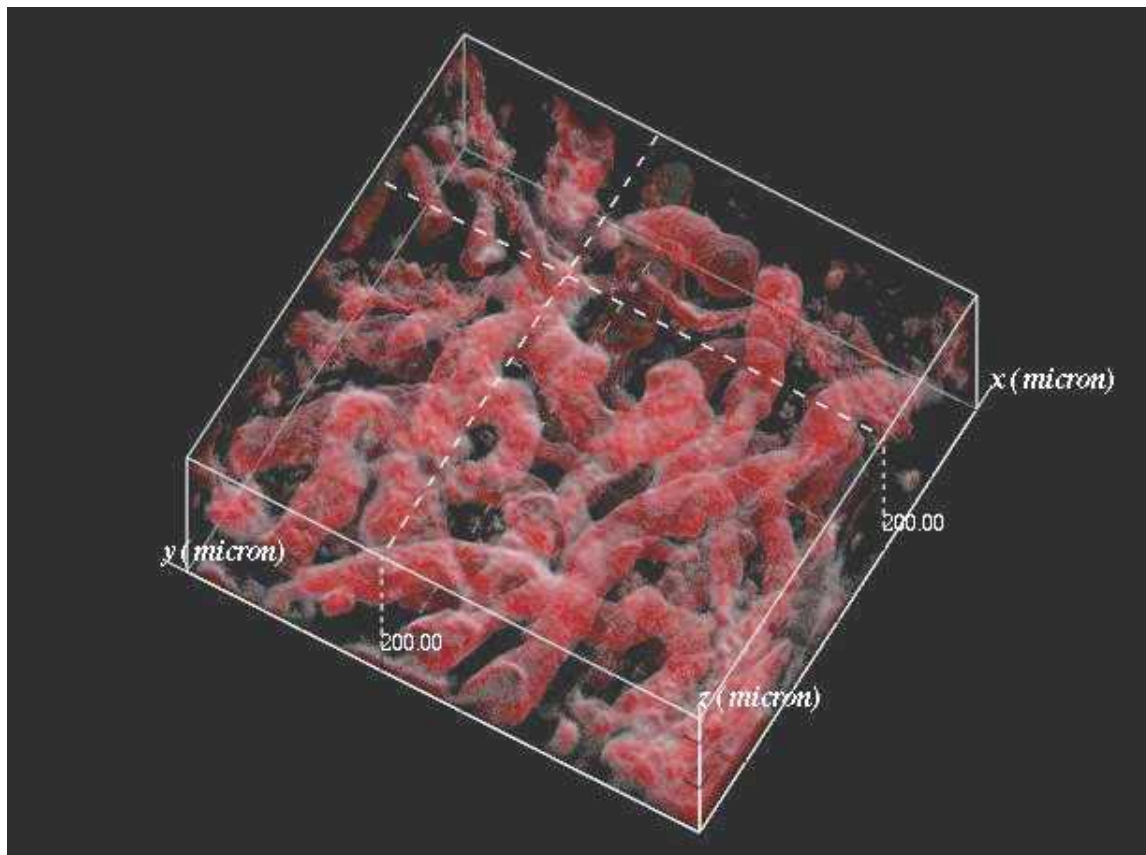


Harvard University Library  
Library Digital Initiative

The Francis A. Countway Library of Medicine  
Harvard School of Public Health, Biomedical Imaging Laboratory

***The Biomedical Image Library (BIL)***



Final Report  
March 12, 2004

## Table of Contents

The Biomedical Image Library: Introduction .....	2
Project Staff .....	2
Scope .....	3
Overview .....	4
Workflow .....	4
<i>Submitting data: the BIL Editor</i> .....	4
<i>Holdings Management</i> .....	6
<i>Using BIL: the BIL user interface and TED</i> .....	6
Data in BIL .....	6
<i>Images</i> .....	6
<i>Metadata and schema</i> .....	8
Development History .....	8
Project Costs .....	10
Future of BIL .....	10
References .....	12
Appendix A: Original BIL Proposal .....	i
Appendix B: Sample Project Record .....	xv

*Cover photomicrograph: A confocal image of a hemangioma in human skin showing tortuous growth of blood vessels. This three-dimensional image was reconstituted from approximately 130 optical sections. The block is approximately 300 microns square and 130 microns tall. BIL project ID 000001. Courtesy of Rick Rogers, Harvard School of Public Health.*

## The Biomedical Image Library: Background

Imaging technology plays a fundamental role in the advancement of biomedical science. The first pictures of cells, the first electron micrographs, and even the first look at stained human chromosomes have all ushered in new ideas about the functioning of living things. Today, with the proliferation of digital technologies, images are more important and pervasive than ever.

Biomedical images appear in periodical scientific literature, either in print or electronic form. Images also appear in atlases, formerly printed but increasingly in electronic formats. Biomedical scientists can find these images only indirectly through traditional indexing of periodical literature (primarily MEDLINE and Biological Abstracts). Furthermore, published images, which only offer a small fraction of the data that underlies the work, are usually offered in low resolution formats such as GIF or as printed images. Both exclude much of the depth of the original digital data, and much data that could be made available for novel analysis remains unavailable. Moreover, digital image data is a management problem for laboratories. Investigators require a straightforward method of archiving image data.

To address the need for large-scale distribution and storage of biomedical image data, we have developed a publicly accessible repository and retrieval system for original digital micrographs, the Biomedical Image Library (BIL). We anticipate that access to images in the original digital format will be of particular benefit to researchers who may wish to reuse existing digital data sets for novel quantitative analysis. BIL also may serve as a repository of data supporting articles in print or electronic publications, allowing scrutiny of data that currently cannot be published economically.

The searchable BIL catalog is available at <http://nrs.harvard.edu/urn-3:hul.eresource:bioimlib>. BIL was funded as a third-round Library Digital Initiative (LDI) grant project awarded in April 2000 (Harvard University Library, 2003).

### Project Staff

#### Countway Library

Paul Bain	BIL Co-Manager
Barbara Callahan	Image Cataloger

#### Harvard School of Public Health

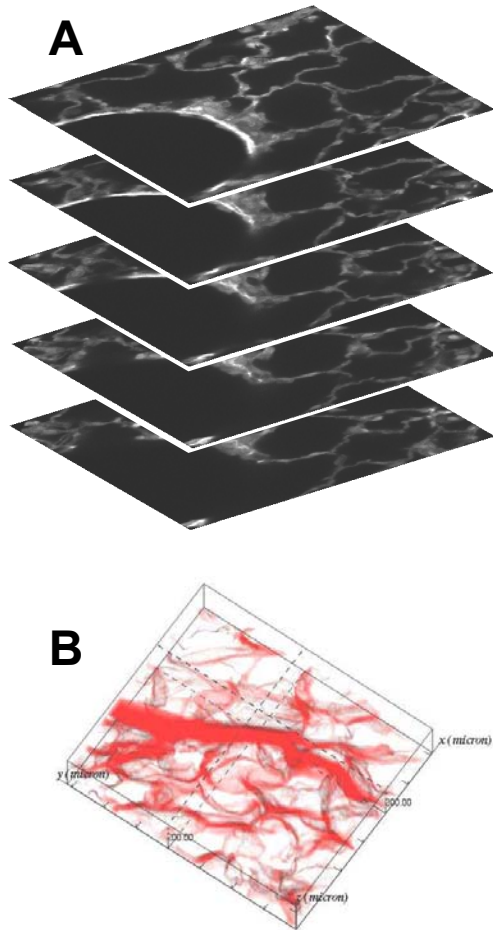
Rick Rogers	BIL Co-Manager
Jean Lai	Image Specialist
Eben Oldmixon	Image Specialist
Claudia Carrera	Student Intern
Agya Soni	Student Intern
Dan Spitsberger	Student Intern

#### Office for Information Systems

Stacy Kowalczyk	OIS Project Manager
Wendy Gogel	Project Liaison
Robin Wendler	Metadata Specialist
Phil Mitchell	Software developer
Stephen Chapman	Preservation Specialist
Ford Fay	Database Developer
Clare McInerney	Database Developer

## Scope

BIL is a repository and retrieval system for original digital photomicrographs produced in support of basic biological research related to human health and development. Users of BIL are the scientists, students, and educators that form the biomedical research community. Biologists, medical



**Figure 1: The image stack.**

Confocal microscopes are capable of capturing multiple frames through a specimen. The resulting image is a stack composed of many individual bit maps (A). Individual frames can be assembled computationally to produce a three-dimensional rendering (B). Images are frequently captured in grey-scale and assigned a false color. In B, blood vessels captured in grey-scale have been assigned a red color.

scientists, and clinicians will use BIL to distribute their work to the community or to identify and retrieve data in support of novel investigations. Access to archived images in the original digital format will be of particular benefit to researchers who may wish to use existing digital data sets in unforeseeable ways. Educators and students will find a ready collection of images to support learning. BIL will also serve as a repository of data supporting articles in print or electronic publications, allowing scrutiny of data that currently cannot be published through traditional means, such as stacks of serial sections recorded from a specimen or moving pictures.

The subject range covered by the images in BIL is intended to be very broad, encompassing the breadth of biomedical research being carried out in the various academic units of Harvard University. Any original digital photomicrographs produced in support of basic biological research related to human health and development can be included in BIL. Images in BIL are available to the public for use in support of scholarly work and education.

Images in BIL are multidimensional representations of biological materials that have been captured through light, fluorescent, or confocal microscopy and that are intended primarily for visual appreciation and study. Modern microscopes capture digital data not just from the surface of a specimen but also at arbitrary depths within it. Image data also can be captured in time and at many specific wavelengths of light. Hence, a biological image is often composed of many two-dimensional frames of view that must be manipulated computationally to produce a final image (Figure 1). The data in BIL are the data captured at the microscope and are not the final, fully rendered illustration that may appear in print.

BIL is intended for representations of biological specimens. Therefore, biomedical images such as electrophoretic gels and their blots that are intended to reveal the composition of samples, rather than spatial relationships in a biological structure, are not included in BIL. Such compositional analyses require significantly different metadata than that accommodated under the current BIL design.

## Overview

As a multi-component system, BIL includes both custom and templated software and makes extensive use of infrastructure provided through the LDI (For details about LDI infrastructure used in BIL, see the sidebar on page this page.). For instance, data are stored and managed in Harvard's Digital Repository Service (DRS), and persistent identifiers for images and associated metadata are issued and managed by the Name Resolution Service (NRS). Likewise, the BIL search interface is generated by the TEmplated Database system (TED), a hosting service that provides an online home for specialized catalogs that do not fit within the scope of existing Harvard catalog systems. Some custom code was developed as well, most significantly the BIL Editor which manages data for investigators in the lab and prepares that data for submission to BIL through other custom components.

## Workflow

### *Submitting data: the BIL Editor*

BIL provides investigators with a tool for organizing images and metadata before submission. The submitter (the investigator with images to contribute), uses a desktop client, the BIL Editor, to create metadata and to send the metadata and images into the BIL workflow (Figure 2, #1-3). The Editor is an interactive, semi-automated system designed to ease the submission process for contributing investigators. Written in Java, its primary purpose is to simplify, as much as is possible, the tasks of creating the intellectual metadata for each of the images in the project. The BIL Editor's interface uses much of the same code base as the TED data maintenance interface system.

The Editor has two main tasks. The first is to construct an instance of metadata in XML that can be loaded into TED. When launched from a computer with access to a local store of image data, the Editor builds a data entry form based on the directory structure of the data store specified by the submitter. Using both the file directory structure and input from the submitter, the software creates a BIL metadata record that abides by the BIL XML schema. The Editor's second task is to package the image data (with the metadata) for submission to the BIL cataloger. Once data entry is complete, the Editor assembles the image data and metadata into a single data file (in tar format) for

### **LDI Infrastructure used in BIL**

#### **DRS (Digital Repository Service)**

*DRS provides Harvard affiliated owners of digital material with a storage and retrieval system for their collections. Digital Repository services and facilities include an electronic storage facility within which the digital objects created or purchased by Harvard agencies reside, management of administrative and structural metadata associated with stored objects, preservation policies and procedures to ensure the continued usability of stored objects, and delivery of an object to a registered or known software application (e.g., an online catalog). For more information about DRS, see <http://hul.harvard.edu/ois/systems/drs/>, Office for Information Systems (2003b).*

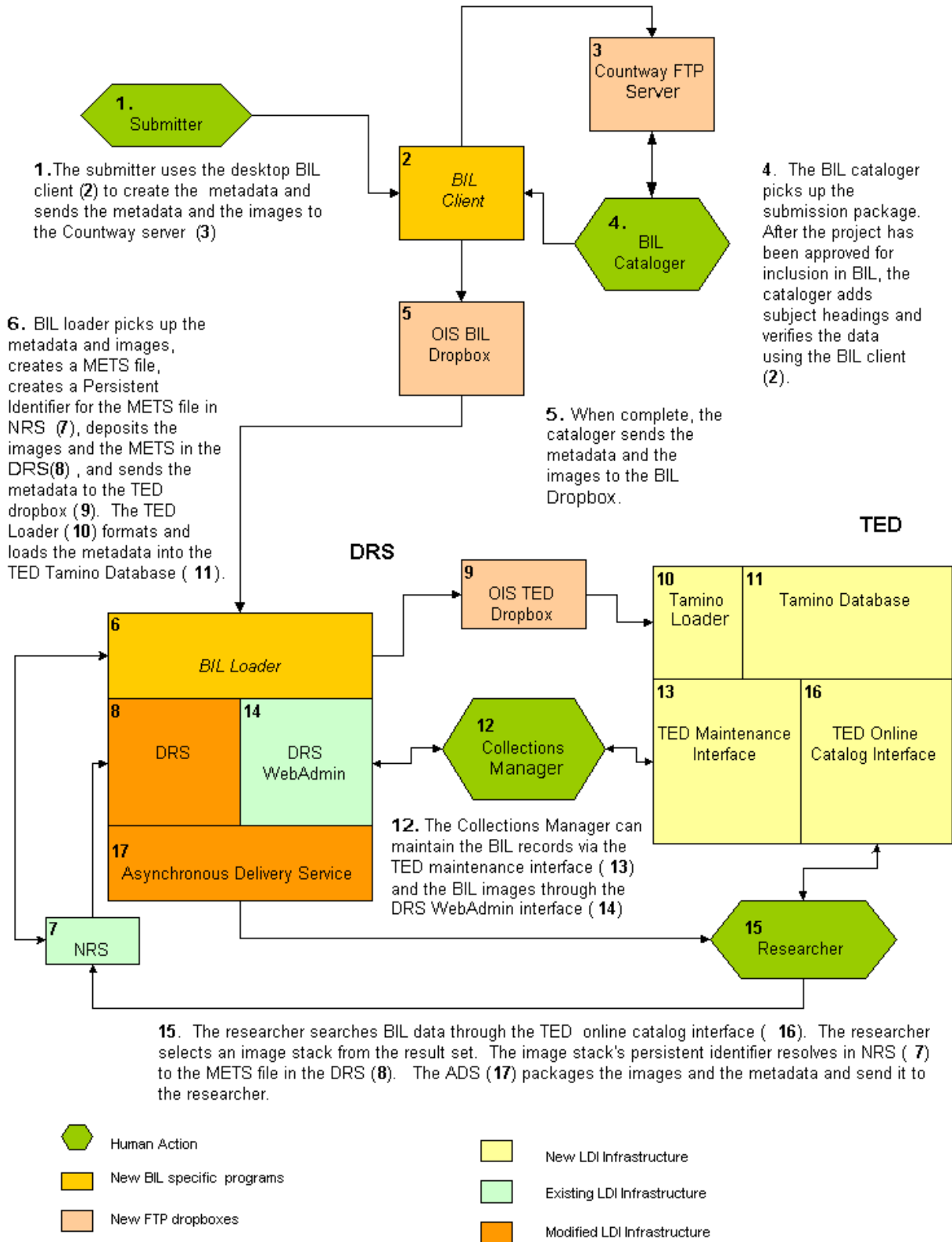
#### **NRS (Name Resolution Service)**

*The Name Resolution Service (NRS) is a comprehensive service for creating, maintaining, and resolving persistent identifiers. A persistent identifier is a permanent, location-independent identifier for a network-accessible resource. Persistent identifiers are sometimes referred to as names. For more information about NRS, see <http://hul.harvard.edu/ois/systems/nrs/>, Office for Information Systems (2003a).*

#### **TED (Templated Database system)**

*Using XML technology, TED allows a curator to create a standard database for a collection, define a structured data format, and easily customize screens and parameters for search and display. This data can be maintained directly in TED or batch loaded from an external source. TED offers a web-based public interface that enables end-users to search, browse, view result sets, and retrieve records for items in a collection, as well as link out to digital objects when such links are provided.*

*BIL was a significant driving force in the decision to create TED because it represented a complex instance of what OIS perceived to be a common request from Harvard Libraries: to construct specialized catalogs for particular collections. OIS undertook the development of TED as a means to simplify and expedite the process of catalog construction. BIL was constructed on top of the TED system by simply customizing the search indexes and appearance of the generic TED interface. For more information about TED, see <http://hul.harvard.edu/ois/systems/ted/>, Office for Information Systems (2003c).*



**Figure 2. Schematic of BIL.** BIL is a multi-component system making use of LDI infrastructure and custom components. See text for further information.

convenient transmission. The Editor then sends this file by FTP to a server at the Countway Library (#2-3).

Compiling and sending data to BIL is the only task that the submitter need complete. Once the data arrive at the Countway, a cataloger verifies the image data and metadata and adds subject headings to images using the National Library of Medicine's Medical Subject Headings (#4). The cataloger forwards completed metadata and the images to a drop box (the OIS BIL Dropbox, #5) where the data await loading into DRS. Loading is carried out at regular intervals by a script called the BIL Loader which retrieves image data and metadata from the dropbox, and creates a METS (Metadata Coding and Transmission Standard) file to store the structural metadata that serves as a manifest and binder for the image data (#6). The loader also names the images and the METS file in NRS for persistent access before it deposits the images into the DRS and the metadata into the TED Tamino database (#7-11). Once in DRS and TED, the images and their metadata become publicly available and part of Harvard's digital collections.

### *Holdings Management*

Librarians at the Countway and at OIS can manage the collections through both the TED maintenance interface, which allows metadata in TED to be viewed and edited, and the DRS WebAdmin interface that provides accounting and administrative-level access to the image files (Figure 1, #12-14).

### *Using BIL: the BIL user interface and TED*

For the user, BIL is a search interface that allows searching and retrieval of image data (Figure 1, #15 and 16). Searching BIL data is carried out through a customized view of the TED online catalog interface. From the BIL interface, users either can search BIL records for specific keywords or can browse BIL's person or subject indexes. Search tools include truncation, wildcards, and limits; all are standard features of TED interfaces.

TED returns the results of the user's search as a list of projects sorted by title. The results display includes the name of the project's principle investigator, the first line of the project abstract, and the date which the project was accepted into BIL. From this point, the user either can choose to save the project as a whole or to view the project record in more detail. From the detailed display, the user can expand or collapse elements to view or hide additional information (Figure 3). At any point, image stacks can be selected and saved for retrieval. A history of previous searches completed during the current session is available.

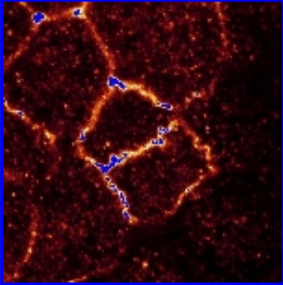
Researchers retrieve image stacks and metadata from a separate screen (#17). From this area, the user can choose to delete material, receive metadata through email, or get image data. Image data is delivered by asynchronous delivery to help the user avoid unexpectedly large downloads. The delivery system packages the image data and metadata in a single zip file for convenient transfer and delivers to the user an URL where the image data can be retrieved.

Help files are available from the BIL online help pages (<http://www.countway.harvard.edu/bil/help.html>) as is more information about the BIL project.

## **Data in BIL**

### *Images*

The images in BIL are all TIFF files. Each image file may be gray scale, color palate or RGB, and the metadata schema has provisions for color maps (look-up tables). As mentioned above,

- Project
  - ◆ + Project Description
  - ◆ + Instrument Setup
  - ◆ - Datasets
    - ◇ - Subgroup
      - **Name:** schneeberger-1881
      - **Description:** Occludin-VSV-G transfected cells, ZO-1 (red) and occludin-VSV-G (green) co-localize at the TJ, producing a discrete yellow signal
      - - **Image Stack**

      - **Created Date:** 1999-07-15
      - **Image Type:** palette color
      - **Planar Configuration:** 1
      - **Creator:** Rogers
      - + **Site**
        - **Subject:** Animal
        - **Subject:** Viral Envelope Proteins
        - **Subject:** Transfection
        - **Subject:** Tight Junctions
      - **Caption:** 1
      - **Rights:** President and Fellows of Harvard University
      - + **Series**
      - + **Preparation**
      - + **Instrument Setup**
        - **Sampling:** purposeful search
        - **Viewing Note:** Occludin-VSV-G transfected cells, ZO-1 (red) and occludin-VSV-G (green) co-localize at the TJ, producing a discrete yellow signal
        - **Artifact:** none identified

**Figure 3. Detailed Display in BIL/TED.** Users can expand (+) or minimize (-) the detailed display of a BIL project record to control the amount of information appearing on the monitor. Here, "Project Description" (top) is minimized to hide the title and abstract, while a "Subgroup" and one "Image Stack" has been expanded to display information about the image whose thumbnail is shown. The boxes (left of "Project" and "Image Stack") can be checked to save data for later retrieval.

current microscope technology allows investigators to create images in more than two dimensions (Figure 1). For instance, a typical confocal micrograph may be composed of 100 single images in depth (the Z deminsion) and in three distinct wavelengths of light (channels). In some experiments, data would be captured in all the above dimensions, but sampled repeatedly over some period of time. Hence, a single image such as the one described above may contain thousands of individual bit maps. In BIL, these multidimensional images are stored as collections of TIFF files referred to as image stacks. Each image stack contains at least one TIFF file and is organized by a METS document (Metadata Encoding and Transmission Standard). By storing images in this way, BIL is able to accommodate complex images of arbitrary structure.

### *Metadata and schema*

In specifying metadata for BIL, our challenge was to design a system that would capture experimental structure in a completely flexible way to avoid forcing contributors to fit their work to our design. At the same time, we wished to simplify the process of changing the metadata needed to accommodate new microscopic technology and practice. XML was an obvious choice, allowing the system to adopt a flexible data structure, while its schema-driven architecture provides a straightforward means of modifying metadata elements. The Tamino-based TED provided an excellent environment in which to implement XML (A sample BIL instance is included as Appendix B, and the schema is available from [http://hul.harvard.edu/ois/xml/xsd/bil/bil\\_schema.xsd](http://hul.harvard.edu/ois/xml/xsd/bil/bil_schema.xsd)). BIL is the first collection to be implemented in TED.

Using XML, BIL project instances are documents that subdivide related images into an arbitrary number of contributor-specified nodes representing the various studies, experiments, treatment groups, and samples. Such a hierarchy allows BIL to preserve the scientific context of each image. It also allows easy browsing of results within the project and enables researchers to retrieve defined, related segments of a project's holdings. At the project level, the schema contains a title and a description of the study, contact information for the investigators in whose laboratories the work was conducted, and references to published worked based on the images. It also includes subject headings from the National Library of Medicine's Medical Subject Headings (MeSH). Information about instrumentation and the procedures used to prepare the biological specimens is included as interchangeable and reusable elements. Hence, information about a microscope or experimental protocol only has to be entered once. Furthermore, adding a new instrument or procedure to the schema, which might require novel data elements, simply involves creating an extension to the original XML schema. These project-level data elements belong to all the subordinate elements in the hierarchy.

Projects contain image stacks. Metadata for each image stack includes a creator, a brief caption describing the image, and a reference to both instrument and preparation information. In addition, each stack can be grouped with other related stacks by creating nodes that correspond to treatment groups, experiments, etc. There is no limit to the number of nodes in the hierarchy, and the names of each node are determined by the contributing investigator. Such a structure allows investigators to control the organization of their own studies and yet provides a uniform scaffold for the image metadata.

### **Development History**

BIL was conceived as a collaboration between the Biomedical Imaging Laboratory at Harvard School of Public Health and the Countway Library to provide researchers a straightforward means of archiving image data. In the late part of 1999, the two groups began preparing a proposal for

consideration by Harvard's Library Digital Initiative. After a preproposal describing a system based on local storage of data received positive reviews by the LDI Grant Review Committee, a final proposal was submitted in March 2000 (Appendix A). The repository described in the final proposal was to be built around storage of image data in the Digital Repository Service. This change was made at the request of the Grant Review Committee with the goal of promoting a more persistent archive. The final proposal was accepted on April 21, 2000 as a three-year project. The start date was set for June 1, 2000.

Initial work focused on defining metadata needs, establishing a file format, and the development of requirements for the BIL system. By August 2000, the team had chosen TIFF as the best format for storing BIL images, and by April 2001, the group was discussing the details of image deposit in DRS, including the creation of the METS structural metadata documents to bind files in stacks. A metadata description for BIL was completed on April 13, 2001 (BIL Metadata Specification). This document, which provided for project and image level description, creator-specified structural levels, and flexible attributes for instrument parameters and procedural details, specified the basic characteristics of BIL metadata and would be the foundation of later XML developments. In early 2001, OIS hired Phil Mitchell to design and code the software necessary for BIL. He and the rest of the team developed a High-level Specification (May 8, 2001) for BIL that includes a review of functional requirements, system architecture and design issues. The specification was vetted in an internal review meeting at the Office for Information Systems on May 17, 2001.

Also in April 2001 the Countway Library hired Barbara Callahan to serve as the Image Cataloger for the project (0.5 fte). Working with Jean Lai and Eben Oldmixon at HSPH, she began to collect metadata for the first two BIL projects.

As OIS developed different implementation models, it became increasingly clear that although a stand-alone application would provide the functionality that BIL required, it would also make it very difficult for both OIS and the Countway staff to maintain and enhance the system over time. The highly structured, deeply nested metadata required a flexible implementation to allow for additional data elements as the types of projects included in BIL grew over time.

At the beginning of July 2001, Phil Mitchell completed a discussion paper for long-term planning. This document analyzed the components of BIL then weighed the advantages and disadvantages of building either a custom system or templated system. Until this point, the team's direction assumed the system would be built as a unique, custom system. Under a templated approach, a generic system would be developed that could be adapted to unique applications such as BIL by XML style sheets and schemas. These development alternatives were discussed at a meeting held on July 12. This discussion marked the birth of the Templated Database (TED) system, though the name had not yet been affixed to the project. The BIL team, along with Dale Flecker and Judy Messerle (Director, Countway Library), decided that TED would be a good base for the BIL catalog and data maintenance interface. The BIL contributor's interface would need to be a small, separate application.

In the second year of development, the decision to proceed with TED was finalized. OIS licensed an XML DBMS (Tamino), and Phil Mitchell developed BIL as an XML schema. The first version of the schema was delivered by February 2002. Since the change to a templated system would mean a six-month delay in BIL's completion, Phil Mitchell's original appointment was extended from December 2002 until the end of the project in June 2003.

During 2002, Dan Spitzberg, a student intern in the Biomedical Imaging Lab, developed an Access database for the temporary storage of BIL metadata and to serve as a test platform. Barbara Callahan completed a metadata dictionary(<http://www.countway.harvard.edu/bil/dictionary.html>) that provides definitions for all of BIL's metadata elements and lists authorities, responsibilities and data formats. She also developed a set of guidelines for use in entering data into any BIL cataloging system. Cataloging of five data sets from the Biomedical Imaging Laboratory was completed.

This period saw the delivery of a beta version of the BIL metadata editor (BIL Editor) for user contributions and the presentation of the TED interface. The BIL Editor was installed on computers in Rick Rogers' laboratory in early May 2002. Early testing resulted in one revision of the software, which was delivered in mid July 2002. The version required further revisions but was adequate for data entry under controlled conditions.

On July 26, 2002, Clare McNerney first demonstrated the TED interface template. At that time, generic search functions were in place. Other features such as object delivery and the clipboard were still in development. Work to customize the interface for BIL began at this time. The Countway/HSPH groups contributed design suggestions, help documentation, and other content for the interface. On March 1, 2003 a finalized version of the BIL/TED interface became publicly available.

Meanwhile, the final version of the BIL Editor was being readied. The final version of the editor to be completed under this project, version 1.0.8, was delivered on June 5, 2003. This version has a number of bug fixes, and in general is both more stable and more flexible than previous versions. Delivery of the final version of the editor marked the end of the development work under this program.

**Project Costs**

Primary funding for the BIL project was an internal grant awarded through LDI. As part of our project proposal, we requested \$364,642 (Appendix A). At the time of approval the

project was granted an upper limit for expenses of \$403,000. Actual charges to the LDI grant amounted to \$319,585.37. Table 1 gives the estimated proportion of funds devoted to distinct work areas. Additional costs, born by all units in supporting BIL, were not tracked closely but are estimated to total approximately \$63,000 across all work areas (Table 1, *Cost Share*).

**Table 1. Cost of BIL.**

Category	LDI Funding	Cost Share*	
		Countway	HSPH
Management and planning	\$Expense	\$Expense (0.25 fte)	
Cataloging	Expense	(ws)	
Image Preparation	Expense		Expense (ws)
Development	Expense		Expense (ws)
	\$320,000	\$57,000	\$6,000

\*Cost share is an estimate of the resources expended by all administrative units in each category. It includes uncompensated salary and fringe benefits for individuals spending greater than 0.1 FTE on the project and the estimated cost of providing workstations (ws) for BIL project use.

**Future of BIL**

BIL was originally conceived as a means of providing storage and retrieval for images produced at Harvard. As proposed, the project specified building a small collection of confocal images. With the initial phase of the project completed, BIL has an inventory of four projects with 834 image stacks represented by jpeg thumbnails. In total there are 2,918 TIFF files in the collection. TED holds for the four project some 417 XML instance docs, representing the total number of nodes in all four projects. The subject matter covered in these images focuses on the growth and development of

blood vessels (angiogenesis) and the response of lung tissue to environmental stress, topics representative of studies being carried out in a large number of Harvard laboratories. All images were collected on confocal microscopes. This limited collection must be regarded as a pilot for the future, which can grow to include much more material and to integrate more smoothly into the daily practice of biomedical investigators.

Adding material to the collection is a priority. Additional projects from the Biomedical Imaging Laboratory are being assembled for deposit into BIL, but it will be necessary to add diversity to the collection. To do so the subject matter of the projects will be expanded to include other areas of research. For instance, we are currently working to add images from the development of the marine acorn worm, an important organism from the point of view of vertebrate evolution. These images, produced as part of a large, genome-scale project, have been produced with a dissecting microscope. Integrating these images will be an excellent test of BIL's extensibility.

Future development of BIL should also include better integration of deposit and use into the routine of laboratory investigation. We have learned that, as BIL is currently configured, contribution to BIL requires a significant investment of time on the part of investigators. Since this time does not directly advance the investigator's research agenda, it represents a barrier to using the system. To encourage the use of BIL, the deposit process either should be transparent or provide some direct benefit to the submitter. Such a system should be capable of recording data directly from microscopes and should provide tools to assist submitters in managing image data locally. Ideally, deposit into a repository should require only a single step such as turning on access or sending the data in a project to BIL. An excellent starting point would be to work with commercial or open-source management and analysis tools such as Open Microscopy Environment (Sorger, Rines, Goldberg, Swedlow, & Hughes, 2004), which are designed to collect and manage image data locally. The goal of this work would be to create a tool that minimizes the time and effort an investigator must spend organizing, publishing, and archiving image data. Making BIL easier to use also will speed the growth of the collection.

We also would like to work with other repositories in sharing access to metadata and holdings, increasing the effective size and value of the holdings available to investigators everywhere. To enable such data sharing, it will be necessary to develop and implement protocols for the exchange of metadata among repositories, perhaps based on Open Archives Initiative protocols. It also will be necessary to develop a minimal metadata description for microscopy images encompassing technical and descriptive elements based on the BIL XML schema or on the very detailed Open Microscopy Environment XML schema.

One possible collaboration is with the BioImage Database at Oxford University (*Bioimage*, 2004). Although its scope is larger than BIL's, BioImage has a well developed collection and infrastructure, and has expressed interest in working with BIL to share holdings information. We are also seeking partner institutions to mount repositories similar to BIL. A network of BIL-like repositories would allow institution to control holdings locally while allowing widespread access.

It is our hope that BIL will become the primary means of distributing biomedical images of current and possibly historic interest from Harvard Medical School and Harvard University. Through collaboration, BIL may also become part of an extensible network for the storage and distribution of digital images among many biomedical research institutions. By improving the direct usefulness to investigators and confederating our holdings with other institutions, we will come closer to these goals.

**References**

- Bioimage*. (2004). Retrieved 23 Feb, 2004, from <http://www.bioimage.org>
- Harvard University Library. (2003, January 2003). *Library Digital Initiative*. Retrieved 2 Feb, 2004, from <http://hul.harvard.edu/ldi/index.html>
- Office for Information Systems. (2003a, 21 October 2003). *Name Resolution Service (NRS)*. Retrieved 23 Feb, 2004, from <http://hul.harvard.edu/ois/systems/nrs/>
- Office for Information Systems. (2003b, 17 September 2003). *Overview: Digital Repository Service (DRS)*. Retrieved 23 Feb, 2004, from <http://hul.harvard.edu/ois/systems/drs/>
- Office for Information Systems. (2003c, 1 April 2003). *Overview: TED (TEmplated Database)*. Retrieved 23 Feb, 2004, from <http://hul.harvard.edu/ois/systems/ted/>
- Sorger, P., Rines, D., Goldberg, I., Swedlow, J., & Hughes, B. (2004, 5 February 2004). *Open Microscopy Environment*. Retrieved 23 Feb, 2004, from <http://www.openmicroscopy.org/>

**Appendix A: Original BIL Proposal, March 17, 2000**

## SECTION A. Grant Proposal

Project unit (Library/Museum/Department): Francis A. Countway Library of Medicine

Project Title: Harvard Biomedical Image Library

## Project Summary:

Imaging technology plays a fundamental role in the advancement of biomedical science. The first pictures of cells, the first electron micrographs, even the first look at stained human chromosomes have all ushered in new ideas about the functioning of living things. Today, with the proliferation of digital technologies, images are more important and pervasive than ever; for example nearly every issue of every biomedical journal contains scientifically informative images. Yet, despite their importance, images cannot be found without navigating the primary literature. Moreover, there is no means for investigators, faculty, or students to retrieve easily large sets of image data. To address these needs, the Francis A. Countway Library of Medicine and the Harvard School of Public Health BioMedical Imaging Laboratory propose to develop a collection of biomedical images, the Biomedical Image Library (BIL). As a system for permanent storage of images and original digital datasets, BIL allows rapid, online storage of high-quality TIFF images (or other open, widely available format) and off-line storage of original datasets in native format. BIL also includes a means for creators (scientists and clinical researchers) to deposit image data and descriptive information into the library. A central, comprehensive catalog will provide an interactive search utility, allowing users to identify images through Medical Subject Headings or through field-specific keywords (author, title, description, etc.). The library will provide investigators with a means to distribute and retrieve image data for review or novel analysis without the need for searching the primary literature or obtain them through the mails. The BioMedical Imaging Laboratory and the Countway Library have produced a demonstration Web site that illustrates some of the ideas presented in this proposal [1].

## Total LDI Requested Funding:

	<u>LDI Funding</u>	<u>Cost Share</u>
Year One Total	\$120,664	salaries + \$3,859
Year Two Total	\$175,318	salaries
Year Three Total	\$68,660	salaries
<b>Project Total</b>	<b>\$364,642</b>	<b>\$75,000</b>

## Project Narrative:

## The Resource

We envision BIL becoming a repository and retrieval system for images produced in support of basic biological research related to human health and development and usually obtained by digitally sampling signals generated by reflected, transmitted or emitted light, radioactive decay, radio frequency resonance, or sound. Images in BIL are two- or three- dimensional representations of biological materials obtained by light, fluorescent, or confocal microscopy and intended primarily for visual appreciation and study. Future BIL acquisitions may include scanning and transmission electron micrographs, nuclear magnetic resonance images, radiographs, or sonograms. BIL images will be restricted to datasets that represent directly the spatial relationships among a sample's

features. This definition excludes many biomedical images that might otherwise be considered, such as images of electrophoretic gels and their blots, as these are intended to reveal the composition of samples, rather than spatial relationships in a biological structure.

As we begin the development of the BIL's infrastructure, we will limit the collection to images created through the digital sampling of transmitted, reflected, or emitted visible light and whose resolution falls between roughly 0.2 and 10 microns (light micrographs). This restriction includes most types of light micrographs, but excludes images produced by many other instruments such as dissecting microscopes, electron microscopes, scanning-tunneling microscopes, and many other imaging technologies. Among the excluded materials are lecture slide sets, and pathology reference materials.

The images we will use to begin development of the library will be assembled from the data archive of the BioMedical Imaging Laboratory. Although the images in this collection exist in a variety of formats, including tiff, rgb, jpeg and gif, they will be archived in a standardized, persistent specification. When possible, we will store images in the format in which the data were captured. We will limit the range of subjects covered by the images, allowing us to build a useful concentration of material within selected domains. One area we will focus on is the growth and development of blood vessels (angiogenesis). This topic, represented by some 8000 records in the BioMedical Imaging Laboratory's archive, has important implications for basic research into cancer and development. It is well represented among the interests of Harvard's research community (a record is one or more images collected in a single imaging session). A second area of concentration will be environmental health. The BioMedical Imaging lab, for instance, has generated about 5000 records documenting the cellular effects of inhaled environmentally important particles fibers on mammalian lung tissue. This area, too, is well represented at Harvard's School of Public Health and is of general interest to the larger biomedical community. We will choose approximately 1000 records of outstanding interest from both areas to provide a foundation for the collection.

After the initial development period, during which time cataloging and storage procedures will have been developed and proven successful, we will expand the scope of the subjects and technologies covered by BIL, adding collections from Harvard laboratories that relate to cancer, regulation of organ development, and environmental health. We also will include digital images produced at submicron resolutions, such as those produced by electron microscopes. By the time we expand BIL to include these additional subjects and technologies we expect to have up to 4,000 images. Once BIL has proven workable for this larger group of images, we will expand the topic coverage again to include images from additional subject areas. At this time, it may also be possible to include a wider range of technologies such as magnetic resonance images or light micrographs with resolutions exceeding 10 microns.

#### Access and Use

Biomedical images appear in periodical literature, either in print or electronic manifestations. Images also appear in atlases, formerly printed but increasingly in electronic formats. Biomedical scientists locate images indirectly through traditional indexing of periodical literature (primarily Medline and Biological Abstracts) or through library catalogs. These images, which only present a small fraction of the data available, are usually offered in formats such as low resolution GIF or printed images that exclude much of the depth of the original digital data. Furthermore, there is no portal through which image data can be made widely available without specific *a priori* knowledge about the existence of that resource.

BIL will be a public, open resource. The users of these images will be the scientists, students, and educators that form the biomedical research community. Biologists, medical scientists, and clinicians will use BIL to distribute their work to the community or to identify and retrieve data in support of novel investigations. Access to archived images in the original digital format will be of particular benefit to researchers who may wish to use existing digital data sets in unforeseeable ways. Educators and students will find a ready collection of images to support learning. BIL will also serve as a repository of data supporting articles in print or electronic publications, allowing scrutiny of data that currently cannot be published through traditional means, such as stacks of serial sections recorded from a specimen or moving pictures.

#### Project Benefit:

BIL will become the primary means of distributing biomedical images of current and possibly historic interest from Harvard Medical School and Harvard University. Although BIL will begin with images located in laboratories at Harvard, we expect that the library will prove to be a valuable asset to investigators around the world. Among the expected benefits of the library are:

- A straightforward means of distributing large volumes of image data to investigators. The library will be able to track what material has been retrieved and by whom, if desired.
- Image data, much of which is unavailable under current publishing practices, will be accessible using BIL's central, comprehensive catalog.
- Educators and students will have easy access to image data for student projects and course work.
- Utilities available within BIL will translate image data from one format to another, simplify access to images regardless of computing platform.
- BIL will be the foundation for an extensible framework for storage and distribution of digital images of medical or biological subjects outside the scope of this application. Such collections could include digital captures of rare medical texts or historically significant photographs.

#### Unique Capabilities and Scope

Although similar data repositories exist, none match the scope or content of BIL. Several clinical image databases exist, either for use within a hospital community or for use as a teaching tool or reference source. For instance, Image Engine, a product of the University of Pittsburgh Center for Biomedical Informatics [2], is a medical informatics system that provides local access to digitized clinical images. A number of electronic atlases have been produced to enhance education and patient care including the Visible Human Project [3,4], The Whole Brain Atlas [5], and the Dermatology Online Atlas [6,7]. Other atlases can be expected to become available at an accelerating rate. None of these sources provides access to research images.

Smaller image libraries focusing on research images are plentiful but typically are narrow in scope, sometimes comprising images from only one or a few laboratories. For example, FlyView is a highly specialized collection of 6500 images showing the expression of anonymous genes in 687 experimental lines of the fruitfly, *Drosophila* [8]. BIL's design as a collection of multi-image projects would provide smaller sites the opportunity to register their collections intact with BIL and also would enable such a resource to be cataloged within BIL with links to each of its images. We will explore methods of accommodating specialized collections within BIL as work progresses. Indeed, such links to holdings outside of BIL suggest a higher-level network of image collections that could make biomedical images located throughout North America or the world available

through a single site on the Web. We consider this an important, long-range goal and shall work to establish collaborations with these other collections.

One project similar to BIL is the European BioImage Database [9,10], which collects multidimensional images of macromolecular structures such as whole virus particles or membranes. Its holdings are distinct from those proposed for BIL in that European BioImage contains views that have been derived from original image data. These images are therefore analytic results or models rather than original data and lie outside the scope of BIL, since BIL's datasets, in contrast, are obtained directly from the sample.

Finally, BIL will have two important advantages over other image projects. One is that it will be designed from the outset to be a unified, central resource for images. Another is that BIL will use universally accepted metadata standards. These two design goals should enable BIL to become an informational hub linking many smaller resources together into an encompassing metaresource.

#### Methodology:

The Biomedical Image Library will allow Web-based access to images and their metadata. Images will be available as a high-fidelity TIFF image (or other standard specification that, like TIFF, can be viewed with contemporary browsers and easily available plug-ins) and in the original acquisition format. A database-driven catalog will store and organize metadata (Fig. 1). Several software suites will manage user access, creator submissions, and administration of the library. BIL's preliminary design calls for a data model based on Dublin Core metadata elements [11], which, with few exceptions, DC elements are applicable to the needs of this project. The LDI's Name Resolution Service (NRS) will assign a unique identifier to each image to ensure persistent access to the holdings.

#### Data Storage and Retrieval

BIL will store image data at three levels.

- 1) The lowest level is a small, low-resolution thumbnail of uniform size that will be associated with the metadata describing each image. Its purpose is to provide each user with a preview of the data he or she can download. These thumbnail images will be stored at the level of the cataloging database. This image will be generated as a part of the cataloging phase from the TIFF images submitted by the creator.
- 2) A high quality image in a standard, persistent specification (probably TIFF) will provide a view of the data adequate for critical examination and printing. The library will provide online access to these TIFF images, which will be stored in the Data Repository Service of the Office for Information Systems. The Office for Information Systems will also manage offline storage. This image will be provided by the creator. Structural information will be added to the metadata fields by the subject analyst and the cataloger.
- 3) A third level will be the data in the format in which the image was originally recorded, which is often proprietary with specialized software requirements. During the first year of the project, original images will be stored online with DRS while we evaluate the usage level and maintenance requirements. After this period, we will decide whether original data should be stored online, offline, or not at all. The library will make no attempt to distribute the software necessary to use these data and will not provide support for viewing or using images stored in proprietary formats.

BIL will deliver all of the files as the creator submits them. The library will neither reprocess images nor conduct additional data analysis. For instance, in the case of images produced by a confocal microscope or other instrument which produces images in multiple planes or optical sections of a specimen, BIL will deliver all the image files corresponding to that specimen and a text file that maps the sequence of the sections. Whenever the creator also produces color-processed files or 3-dimensional constructs from the sequence of files as part of the research, these will be offered for delivery by BIL also. However, BIL staff will not engage in the labor-intensive process of creating new, customized constructs or color files that were not part of the files originally contributed to the collection.

### User Access

Users will access images through a Web-based interface generated by the Biomedical Images Library's access software (Fig. 1, #1). Queries will be made from the interface to the database that stores descriptive information for objects in BIL. Users will have the option of searching by keywords drawn from selected indexes (title, creator, affiliation, methodology, etc.) or by subject using the National Library of Medicine's robust and extensive Medical Subject Headings (MeSH). In our preliminary work, MeSH has proved adequate to describe any biomedical-related image, and we anticipate that the thesaurus will provide excellent access to BIL's resources. We will provide a help function for using MeSH. One further capability of the search interface is the ability to find related images once a user has identified a relevant image. We will provide a tracing function to allow users to identify related images.

Biomedical datasets such as images do not exist in isolation and are always related to other images as part of a study or experimental project. For instance, a study using a fluorescent antibody to examine a line of cultured cells will contain several images, including one or more control images and several images of treated cells. An important feature of the catalog, therefore, will be to organize images into groups that reflect the intellectual structure of the information. We call these groups projects. For this library, projects are distinct objects that bind together a group of related images. They are made up of unique, project-specific metadata elements including a separate title, creator list, and description of the project. Projects will also be assigned MeSH terms that capture the facets of all the images in the project. This logical structure will allow users to retrieve all scientifically related images under a single object.

Once a user initiates a search, the access software will generate a results set that will present a list of relevant projects and images (Fig. 1, #2). The user will be asked to refine the search using an interactive menu. Users will be able to search projects or images or both; results will be presented as a list of projects, and if a listing of images is desired, they will appear as thumbnails with title and creators listed under their parent project. By displaying the images along with their projects we insure that images are never lost from their experimental context. Links from the results set will contain instructions for downloading either a high-fidelity TIFF image or the original data file from DRS using the NRS (Fig. 1, #3). The access software will include help information describing the use of BIL and the identity of specialized software required for those images in original format.

### Creator Submissions

Participating laboratories and investigators will submit images and descriptive information through a standard interface for inclusion into the collection (Fig. 1, #4). A forms-based interface will guide investigators in entering accurate and relevant information to assist in the cataloging of each project and its constituent images. Descriptive information will be deposited directly into the database, but records will not be available until after they have been reviewed and edited by the library's cataloger. Submitted image data will be deposited in a drop-off box for temporary storage

(Fig. 1, #5). For submissions from creators other than the BioMedical Imaging Laboratory, project managers and/or members of the advisory committee will determine whether submitted material is suitable for inclusion into the collection. Images from the BioMedical Imaging will be selected by Rick Rogers, the subject analyst and other HSPH staff. Once the decision to accept an object has been made, the cataloger will proceed to describe the object. Once cataloging is completed, the image data will be moved to the DRS. Creators will be responsible for submitting the TIFF images in the correct format; the library will use this image to produce a thumbnail at a standard size and resolution.

### Administrative Software

Specialized administrative interfaces will allow cataloging, access control, tracking, and reporting on the use of the collection (Fig. 1, #6). The access software will keep track of users and downloads, reporting these statistics to the database. These data will be accessible by the appropriate administrator. The submission software will notify the cataloger and acquisitions specialist of new submissions.

### Staffing

Two work groups, one from the Countway Library and one from the Harvard School of Public Health's BioMedical Imaging Laboratory, will develop BIL in consultation with the Office for Information Systems. The project will be co-managed by Paul Bain (Countway, 0.5 FTE) and Rick Rogers (HSPH, 0.2 FTE). Paul Bain will serve as the contact point for communication between OIS and the BIL management. He will be responsible for tracking the progress of the project, submitting reports, and coordinating the activities of all the parties involved in this project. Rick Rogers will manage the extraction of data from his archive and will be responsible for soliciting contributions from other creators and units within the Medical Area.

At the Countway Library, a cataloger (Stephanie King, 0.5 FTE) will describe images as they are added to the collection and will work with the rest of the project team in the selection of metadata elements, the design of the database, and the development of the cataloger's interface.

The team at the Harvard School of Public Health will consist of an image specialist (Jean Lai, 0.4 FTE), a subject analyst (Eben Oldmixon, 0.2 FTE), and a research intern (Agya Soni, 0.4 FTE, summers only; work will be timed to coincide with her availability). The image specialist will sort and retrieve offline image data from DAT tapes in the BioMedical Imaging Laboratory's archive. The image specialist will then consolidate image files into projects and select pilot images that will be used to develop cataloging rules and procedures. The subject analyst (Eben Oldmixon, 0.2 FTE) will assist the image specialist in selecting and consolidating the image files. The subject analyst will also identify and provide structural metadata based on his scientific background and knowledge for each image and will work with the cataloger in providing descriptive metadata for the projects and images. The research intern (Agya Soni, 0.4 FTE, summers only. Work will be timed to coincide with her availability) will assist in the description of the images and help organize material extracted from the archive.

OIS will develop the BIL catalog including the database, access software, submission software, and administrative interfaces. OIS will also provide consulting services in choosing standard data specifications, and the Countway/School of Public Health group will work with the OIS metadata analyst to refine the metadata requirements of the library. BIL will make use of Data Repository Service and the Naming Resolution Service.

Time Line: Years 1, 2, 3 and Beyond:

The first two years of the project will see the development of the institutional framework and expertise to acquire, catalog, store, and distribute digital image holdings. The actual material contained in BIL at the end of these first two years is expected to lack the breadth of subject matter that is to be expected of the fully-developed BIL.

The third year, therefore, is crucial, as this may be seen as the evangelical period, when BIL actively seeks participants from Harvard, first, then from sister academic and research institutions in North America, and then from abroad. By the end of the third year, BIL should have earned a respected presence as the principal repository of original biomedical digital images on the WorldWide Web.

Enthusiasm for the long-term operation of BIL seems to be assured; the Countway Library is committed to providing support for the storage of objects collected during the development of this project. The Countway will also solicit funding for staffing and maintaining the library into the future. The Countway will also provide user support for the library. Dr. Rogers will work with the Faculty at the Harvard School of Public Health, federal, and private sources for long-term support.

The ongoing operations of the mature library will be overseen by an advisory committee composed of biologists, image technologists, or clinicians who create and use image data. The committee will establish an acquisitions policy, direct expansion of the library's scope, and guide decisions concerning technological changes. The committee will be convened early in the development of the library.

#### Intellectual Property and Rights:

Creators will transfer limited rights to distribute their unpublished materials to the Biomedical Image Library upon their submission. We will require that creators submit a signed statement declaring that limited rights has been transferred to the library. For published images, we will ask the contributor to seek the permission of the copyright holder before these images can be distributed through the library. If permission cannot be obtained, we will not distribute the image. Working with the Harvard Office of the General Counsel (Allan A. Ryan, University Attorney), we have developed a statement that will appear on the entry page to the BIL collection [12]. This statement authorizes users to download and use images as needed in the course of scholarly research, including publication. We will request that users credit the source of the image data and will provide accurate citation information.

#### Long-term Commitment:

The Countway is committed to providing support for the long-term storage and bibliographic maintenance of objects collected during the development of this project. The Countway will also provide research and reference services for these materials. The Countway will seek future sources of funding for staffing, maintaining and expanding the library.

#### **User Input**

We will solicit user input at all stages of the project. As we begin design of the user interface, for example, we will convene a focus group of microscopists to investigate the best methods for accessing the database (see workplan). Once initial development of the user and submission interfaces are complete, we will conduct tests with uninitiated users from the scientific community to evaluate the quality of recall and ease of use of the controls and search functions as they are implemented in the user interface. In the libraries mature form, user and creators will be able to evaluate the interfaces to the library through on-line feedback forms (Fig. 2). Information gleaned

from these forms will be particularly useful in the early months of BIL's public operation (year three) when users are confronted with novel ways of retrieving and submitting familiar materials. An important measure of success for the library will be its impact on research and scholarly communication. Although objectively measuring impact is difficult, usage statistics such as the number of objects retrieved or number of searches conducted will give some indication of who is using the resource and how often it is used.

Only with a clear connection to its users will the library live up to its full potential. Thus, we intend BIL to function within the scientific community. An advisory committee made up of active scientists, potential creators, and users of the material in BIL, will ensure that the ongoing needs of the community are being met. The committee will be assembled during the first year of the project to oversee development and will continue to function after the library is mature, guiding acquisitions policy and technological developments in interface design and storage methods. We anticipate twice-yearly meetings of the committee (see Section B: Workplan). We also will gather information informally through conversations with colleagues and scientists in the biomedical community. These conversations will be useful in evaluating the potential worth and impact of the library on the community and in generating novel uses for the resource.

## References

- [1] Biomedical Image Library (<http://www.hsph.harvard.edu/physiology/projects/bil/>). Francis A. Countway Library of Medicine. (Accessed 3/3/2000). **The username for entry to this site is "imagelibrary"; the password is "bil".**
- [2] Image Engine Homepage (<http://www.cml.upmc.edu/cml/imageengine/imageEngine.html>). Clinical Multimedia Laboratory, University of Pittsburgh Center for Biomedical Informatics. (Accessed 3/3/1999).
- [3] Ackerman, M.J. The Visible Human Project: a resource for education. *Academic Medicine* **74**:667-670 (1999).
- [4] The Visible Human Project ([http://www.nlm.nih.gov/research/visible/visible\\_human.html](http://www.nlm.nih.gov/research/visible/visible_human.html)). National Library of Medicine (Accessed 3/3/2000).
- [5] The Whole Brain Atlas (<http://www.med.harvard.edu/AANLIB/home.html>). Johnson, K.A. and Becker, J.A. (Accessed 10/11/1999).
- [6] Diepgen, T.L. and Eysenbach, G. Digital images in dermatology and the Dermatology Online Atlas on the World Wide Web. *Journal of Dermatology* **25**:782-787 (1998).
- [7] Dermatology Online Atlas ([http://www.dermis.net/bilddb/index\\_e.htm](http://www.dermis.net/bilddb/index_e.htm)). Friedrich-Alexander-University of Erlangen-Nürnberg, School of Medicine, Department of Dermatology (Accessed 3/3/2000).
- [8] Flyview: a Drosophila image database (<http://flyview.uni-muenster.de/Home.html>). Westfälische Wilhelms-Universität Münster, Institut für Allgemeine Zoologie und Genetik, Arbeitsgruppe Entwicklungsgenetik. (Accessed 3/3/2000).
- [9] Carazo, J.M. and Stelzer, E.H. The BioImage Database Project: organizing multidimensional biological images in an object-relational database. *Journal of Structural Biology* **125**:97-102 (1999).

- [10] The BioImage Homepage (<http://www-embl.bioimage.org/>). The BioImage Database Consortium (Accessed 3/3/2000).
- [11] Dublin Core Metadata Initiative (<http://purl.org/DC>). (Accessed 02/16/2000).
- [12] Such a statement might read: Copyright 2000 President and Fellows of Harvard College. Authorized users of this holding may download and use images as needed in the course of scholarly research, including publication of research. We request that you credit the source of the image data as "Biomedical Image Library". For all other permissions, contact [\\_](#).

SECTION B. Draft Workplan

Note: This plan of work assumes bi-monthly meetings with the Library Digital Initiative project liaison to assess progress and to address problems as they arise. Regular reports will be forwarded to the BIL advisory committee.

**Year One**

<b>Year One Goals</b>	
1.	Organize and store angiogenesis and environmental particle images at Data Repository Service.
2.	Establish metadata needs of the library.
3.	Begin systems and database development.
4.	Establish advisory committee and establish acquisitions policy

	<u>Task</u> (tracked by number and letter; 1.A., etc.)	<u>Responsibility</u>	<u>Finish Date</u>
<b>Month 1</b>	<b>1. Organize images in the BioMedical Image Laboratories collection</b>		<b>Month 10</b>
	A. Select and retrieve images from the BioMedical Imaging Laboratories collection.	HSPH	Month 6
	• Search the archive		
	• Retrieve suitable images		
	• Consolidate the image files and organize into projects		
	• Select pilot images	Countway/HSPH	Month 3
	B. Define image format specifications	/ LDI	
			<b>Month 8</b>
	<b>2. Establish metadata needs and develop cataloging procedures</b>		
	A. Identify metadata elements needed to describe images and projects	Countway/LDI	Month 4
<b>Month 2</b>	<b>3. System development</b>		<b>Month 18</b>
	A. Begin scoping and planning database and	Countway/LDI	Month 5

	<u>Task</u> (tracked by number and letter; 1.A., etc.)	<u>Responsibility</u>	<u>Finish Date</u>
	system <ul style="list-style-type: none"> <li>• Database</li> <li>• Access software</li> <li>• Submission software</li> <li>• Cataloging software</li> <li>• Administrative software</li> </ul>		
<b>Month 3</b>	<b>2. Establish metadata needs and develop cataloging procedures</b>		<b>Month 8</b>
	B. Begin describing pilot images and projects	Countway/HSPH / LDI	Month 6
	C. Develop documentation for cataloging rules, specifications and procedures	Countway/LDI	Ongoing, but see month 24
	<b>4. Convene advisory committee</b>		Ongoing;
	A. Work to establish acquisitions policy, etc.	HSPH/Countway	twice a year
<b>Month 4</b>	<b>5 Deposit images with Data Repository Service</b>		<b>Month 7</b>
	A. Establish naming conventions; register with NRS	Countway/LDI	Month 6
	B. XML batch documentation	LDI	Month 6
	C. Complete assessment by preservation department	LDI	Month 6
<b>Month 5</b>	no tasks beginning		
<b>Month 6</b>	<b>2. Establish metadata needs and develop cataloging procedures</b>		<b>Month 8</b>
	D. Test access to pilot images and projects	Countway/HSPH	Month 8
	E. Modify metadata if needed and revise guidelines	Countway/LDI	Month 8
	<b>3. System development</b>	LDI	<b>Month 18</b>
	B. Develop and implement database and systems <ul style="list-style-type: none"> <li>• Database</li> <li>• Access software</li> <li>• Submission software</li> <li>• Cataloging software</li> <li>• Administrative software</li> </ul>		Month 16
<b>Month 7</b>	<b>1. Organize images in the BioMedical Image Laboratories collection</b>		<b>Month 10</b>
	C. Produce thumbnails and TIFF images from each image. Add structural metadata to TIFFs.	HSPH	Month 10
	D. Identify software necessary for viewing each native image	HSPH	Month 10

	<u>Task (tracked by number and letter; 1.A., etc.)</u>	<u>Responsibility</u>	<u>Finish Date</u>
	<b>5. Deposit images with Data Repository Service</b>		<b>Month 7</b>
	D. Transfer images to DRS from Biomedical Image Laboratory	HSPH/LDI	Month 7
	<b>6. Begin subject analysis and description of images and projects</b>	Countway/HSPH	<b>Month 15</b>
	<b>Finish Task 5</b>		
<b>Month 8</b>	<b>Finish Task 2</b>		
<b>Month 9</b>	<b>4. Convene advisory committee</b>		
	A. Establish acquisitions policy, etc.		
<b>Month 10</b>	<b>Finish Task 1</b>		
<b>Month 11</b>	no tasks beginning		
<b>Month 12</b>	no tasks beginning		

Year Two

<b>Year Two Goals</b>	
1.	Begin accepting submissions from other Harvard units.
2.	Complete development of systems and database.
4.	Open access to BIL to the public.
5.	Complete planning for the long-term administration of BIL.
6.	Seek continued funding for BIL.

	<u>Task</u>	<u>Responsibility</u>	<u>Finish Date</u>
<b>Month 13</b>	<b>7. Add materials from other Harvard Units</b>		<b>Ongoing</b>
	A. Invite submissions from other Harvard units	HSPH/Countway	
	<b>8. Draft requirements for the ongoing administration of the library</b>	Countway/HSPH/ LDI	<b>Month 24</b>
<b>Month 14</b>	no tasks beginning		
<b>Month 15</b>	<b>7. Add materials from other Harvard Units</b>		<b>Ongoing</b>
	B. Begin adding and cataloging materials from other Harvard units	Countway/HSPH	

	<u>Task</u>	<u>Responsibility</u>	<u>Finish Date</u>
	<b>4. Convene advisory committee</b>		
	<b>Finish Task 6</b>		
<b>Month 16</b>	<b>3. System development</b>		<b>Month 18</b>
	C. Test and modify initial designs	LDI	Month 18
	• Database		
	• Access software		
	• Submission software		
	• Cataloging software		
	• Administrative software		
<b>Month 17</b>	<b>9. Open BIL for public access</b>		
<b>Month 18</b>	<b>Finish task 3</b>		
<b>Month 19</b>	no tasks beginning		
<b>Month 20</b>	no tasks beginning		
<b>Month 21</b>	<b>4. Convene advisory committee</b>		
<b>Month 22</b>	no tasks beginning		
<b>Month 23</b>	<b>10. Invite submission from units outside of Harvard</b>	Countway/HSP H	<b>Ongoing</b>
<b>Month 24</b>	<b>8. Finish requirements for the ongoing administration of the library</b>	Countway/HSP H/ LDI	
	<b>Also - all cataloging rules and procedures in place and documented</b>		

Year Three

<b>Year Three Goals</b>
1. Continue to add materials from other Harvard units.
2. Open BIL for submissions from outside of Harvard.
3. Expand scope to include electron micrographs and low magnification light micrographs.
3. Evaluate and improve all system, user feedback, help functions, etc.
4. Obtain continuing funding.

Task Responsibility Finish Date

	<u>Task</u>	<u>Responsibility</u>	<u>Finish Date</u>
<b>Month 1</b>	<b>11. Begin adding submission from outside Harvard to the library</b>	Countway/HSP H	<b>Ongoing</b>
<b>Month 2</b>	<b>12. Begin adding electron micrographs and other technologies</b> A. Identify pilot images and begin evaluating metadata requirements	Countway/HSP H	<b>Month 4</b>
<b>Month 3</b>	<b>4. Convene advisory committee</b>		
<b>Month 4</b>	<b>12. Begin adding electron micrographs and other technologies</b> B. Begin cataloging pilot electron micrographs	Countway/HSP H	<b>Month 6</b>
<b>Month 5</b>	<b>13. Evaluate all systems and identify areas for improvement</b>	Countway/HSP H/ LDI	<b>Month 7</b>
<b>Month 6</b>	<b>12. Begin adding electron micrographs and other technologies</b> C. Begin adding electron micrographs to collection	Countway/HSP H	
<b>Month 7</b>	<b>14. Make changes as identified in task 13</b>	Countway/HSP H/ LDI	<b>Month 10</b>
<b>Month 8</b>	no tasks beginning		
<b>Month 9</b>	<b>4. Convene advisory committee</b>		
<b>Month 10</b>	no tasks beginning		
<b>Month 11</b>	no tasks beginning		
<b>Month 12</b>	no tasks beginning		

## SECTION C. Budget

	<b>LDI Funding</b>	<b>Cost Share</b>
Year One		
<u>Salaries (incl. fringe)</u>		
1 Project Manager @ 50 % FTE		50% full-time salary
1 Project Manager @ 20 % FTE	\$LDI Funds	

1 Cataloger @ 50% FTE	\$LDI Funds	
1 Assistant Project Manager 5% FTE		5% full-time salary
1 Systems Liaison @ 20% FTE	\$LDI Funds	
1 Image Specialist @ 40% FTE	\$LDI Funds	
1 Summer Intern @ 15 hrs/wk 14 wks	\$LDI Funds	
1 Programmer @ 6 mo., \$50/hr	\$LDI Funds	

Equipment (incl. Hardware/Software)

Cataloging Workstation		\$2,000
40 Gb Hard drive (for image cache)		880
DAT drive		570
25 DAT tapes		150
Adobe Photoshop		259

Preservation

Consultation/Assessment	\$0.00	
-------------------------	--------	--

Year One Total	\$120,664	salaries + \$3,859
----------------	-----------	--------------------

Year Two

Salaries (incl. fringe and 4% cost of living increase)

1 Project Manager @ 20 % FTE		20% full-time salary
1 Project Manager @ 20 % FTE	\$LDI Funds	
1 Cataloger @ 50% % FTE	\$LDI Funds	
1 Assistant Project Manager @ 5% FTE		5% full-time salary
1 Systems Liaison @ 20% FTE	\$LDI Funds	
1 Image Specialist @ 40% FTE	\$LDI Funds	
1 Summer Intern @ 15 hrs/wk 14 wks	\$LDI Funds	
1 Programmer @ 12 mo., \$50/hr	\$LDI Funds	

Year Two Total	\$175,318	salaries
----------------	-----------	----------

Year Three

Salaries (incl. fringe and 4% cost of living increase)

1 Project Manager @ 20 % FTE		20% full-time salary
1 Project Manager @ 20 % FTE	\$LDI Funds	
1 Cataloger @ 50% % FTE	\$LDI Funds	
1 Assistant Project Manager 5% FTE		5% full-time salary
1 Systems Liaison @ 20% FTE	\$LDI Funds	
1 Image Specialist @ 20% FTE	\$LDI Funds	
1 Summer Intern @ 15 hrs/wk 14 wks	\$LDI Funds	

Year Three Total	\$68,660	salaries
------------------	----------	----------

<b>Project Total</b>	<b>\$364,642</b>	<b>\$75,000</b>
----------------------	------------------	-----------------

## Appendix B: Sample Project Record

The following XML document is an actual BIL project record (project ID 000001), Cellular markers that distinguish the phases of hemangioma during infancy and childhood, from the Biomedical Image Laboratory. Only one image stack has been included.

```
<?xml version="1.0" encoding="UTF-8"?>
<Project xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="http://hul.harvard.edu/ois/xml/xsd/bil/bil_schema-
prod.xsd">
<Project_Description projectID="000001" Schema_Version="1.0"
Submitted_Date="2002-11-15" Last_Modified_Date="2002-12-04">
  <Title>Cellular markers that distinguish the phases of hemangioma during
infancy and childhood.</Title>
  <Abstract>Hemangiomas, localized tumors of blood vessels, appear in
approximately 10-12% of Caucasian infants. These lesions are characterized by a
rapid proliferation of capillaries for the first year (proliferating phase),
followed by slow, inevitable, regression of the tumor over the ensuing 1-5 yr
(involuting phase), and continual improvement until 6-12 yr of age (involuting
phase). To delineate the clinically observed growth phases of hemangiomas at a
cellular level, we undertook an immunohistochemical analysis using nine
independent markers. The proliferating phase was defined by high expression of
proliferating cell nuclear antigen, type IV collagenase, and vascular endothelial
growth factor. Elevated expression of the tissue inhibitor of metalloproteinase,
TIMP 1, an inhibitor of new blood vessel formation, was observed exclusively in
the involuting phase. High expression of basic fibroblast growth factor (bFGF)
and urokinase was present in the proliferating and involuting phases. There was
coexpression of bFGF and endothelial phenotypic markers CD31 and von Willebrand
factor in the proliferating phase. These results provide an objective basis for
staging hemangiomas and may be used to evaluate pharmacological agents, such as
corticosteroids and interferon alfa-2a, which accelerate regression of
hemangiomas. By contrast, vascular malformations do not express proliferating
cell nuclear antigen, vascular endothelial growth factor, bFGF, type IV
collagenase, and urokinase. These data demonstrate immunohistochemical
differences between proliferating hemangiomas and vascular malformations which
reflect the biological distinctions between these vascular lesions.</Abstract>
  <Subject>Adolescence</Subject>
  <Subject>Tumor Markers, Biological</Subject>
  <Subject>Tissue Inhibitor of Metalloproteinases</Subject>
  <Subject>Proliferating Cell Nuclear Antigen</Subject>
  <Subject>Nuclear Proteins</Subject>
  <Subject>Neoplasm Staging</Subject>
  <Subject>Mast Cells</Subject>
  <Subject>Lymphokines</Subject>
  <Subject>Infant, Newborn</Subject>
  <Subject>Infant</Subject>
  <Subject>Human</Subject>
  <Subject>Hemangioma</Subject>
  <Subject>Glycoproteins</Subject>
  <Subject>Gelatinase B</Subject>
  <Subject>Fibroblast Growth Factor 2</Subject>
  <Subject>Endothelial Growth Factors</Subject>
  <Subject>Collagenase</Subject>
  <Subject>Child, Preschool</Subject>
  <Subject>Child</Subject>
  <Support Title="AIDS ppg" ID="MOIRR02172" Funding_Organization="National
Institutes of Health"></Support>
</Project_Description>
<Site ID="1" Site_Identifier="Children's Hematology/Oncology">
  <Department>Division of Hematology/Oncology</Department>
  <Institution>Children's Hospital</Institution>
  <Address_Line_1>Enders 614</Address_Line_1>
  <Address_Line_2>300 Longwood Avenue</Address_Line_2>
```

```

<City>Boston</City>
<State>MA</State>
<Postal_Code>02115</Postal_Code>
<Country>USA</Country>
</Site>
<Site ID="6" Site_Identifier="HSPH">
  <Department>Department of Enviromental Health</Department>
  <Institution>Harvard School of Public Health</Institution>
  <Address_Line_1>665 Huntington Avenue</Address_Line_1>
  <City>Boston</City>
  <State>MA</State>
  <Postal_Code>02115</Postal_Code>
  <Country>USA</Country>
</Site>
<Site ID="5" Site_Identifier="Children's Surgery">
  <Department>Department of Surgery</Department>
  <Institution>Children's Hospital</Institution>
  <Address_Line_1>300 Longwood Avenue</Address_Line_1>
  <City>Boston</City>
  <State>MA</State>
  <Postal_Code>02115</Postal_Code>
  <Country>USA</Country>
</Site>
<Site ID="4" Site_Identifier="Children's Pathology/Pediatrics">
  <Department>Departments of Pathology and Pediatrics</Department>
  <Institution>Children's Hospital</Institution>
  <Address_Line_1>300 Longwood Avenue</Address_Line_1>
  <City>Boston</City>
  <State>MA</State>
  <Postal_Code>02115</Postal_Code>
  <Country>USA</Country>
</Site>
<Site ID="3" Site_Identifier="HMS">
  <Department>Departments of Surgery and Anatomy and Cellular
Biology</Department>
  <Institution>Harvard Medical School</Institution>
  <Address_Line_1>240 Longwood Avenue</Address_Line_1>
  <City>Boston</City>
  <State>MA</State>
  <Postal_Code>02115</Postal_Code>
  <Country>USA</Country>
</Site>
<Site ID="2" Site_Identifier="Children's Plastic Surgery">
  <Department>Division of Plastic Surgery</Department>
  <Institution>Children's Hospital</Institution>
  <Address_Line_1>300 Longwood Avenue</Address_Line_1>
  <City>Boston</City>
  <State>MA</State>
  <Postal_Code>02115</Postal_Code>
  <Country>USA</Country>
</Site>
<Principal_Investigator>
  <Name>Ezekowitz RAB</Name>
  <Email>Unknown</Email>
  <SiteRef ID="3" Name="HMS"></SiteRef>
</Principal_Investigator>
<Co-Investigator>
  <Name>Takahashi K</Name>
  <Email>Unknown</Email>
  <SiteRef ID="2" Name="Children's Plastic Surgery"></SiteRef>
</Co-Investigator>
<Co-Investigator>
  <Name>Mulliken JB</Name>
  <Email>Unknown</Email>

```

```

    <SiteRef ID="1" Name="Children's Hematology/Oncology"></SiteRef>
  </Co-Investigator>
  <Co-Investigator>
    <Name>Kozakewich HPW</Name>
    <Email>Unknown</Email>
    <SiteRef ID="4" Name="Children's Pathology/Pediatrics"></SiteRef>
  </Co-Investigator>
  <Co-Investigator>
    <Name>Folkman J</Name>
    <Email>Unknown</Email>
    <SiteRef ID="5" Name="Children's Surgery"></SiteRef>
  </Co-Investigator>
  <Co-Investigator>
    <Name>Rogers RA</Name>
    <Email>rogers@bioimage.harvard.edu</Email>
    <SiteRef ID="6" Name="HSPH"></SiteRef>
  </Co-Investigator>
  <Publication Author="Takahashi K" Title="Cellular markers that distinguish the
  phases of hemangioma during infancy and childhood." Journal="J Clin Invest"
  Volume="93" Issue="6" Pages="2357-2364" Year="1994" PMID="7911127"
  Usage="original"></Publication>
  <Preparation ID="1" Preparation_Identifier="Single channel 488">
    <Fluorescence_Microscopy>
      <Fixative>4% paraformaldehyde</Fixative>
      <Embedment>spurr plastic</Embedment>
      <Mount_Medium>None</Mount_Medium>
      <Sectioning>trimmed biopsy pieces</Sectioning>
      <Staining>
        <Antibody>None</Antibody>
        <Antigen>None</Antigen>
        <Fluorescent_Stain>lucifer yellow</Fluorescent_Stain>
        <Counterstain>None</Counterstain>
      </Staining>
    </Fluorescence_Microscopy>
  </Preparation>
  <Instrument_Setup ID="1" Setup_Identifier="Confocal">
    <Confocal_Microscope>
      <Manufacturer>Molecular Dynamics</Manufacturer>
      <Model>Sarastro 2000 CLSM</Model>
      <Number_Of_Channels>1</Number_Of_Channels>
      <Channel_Description Channel_Number="1" Optical_Mode="Fluorescence"
  Incident_Light="488 nanometers" Detected_Light=">510
  nanometers"></Channel_Description>
      <Objective_Medium>oil</Objective_Medium>
      <Objective_Magnification>20x, 40x</Objective_Magnification>
    </Confocal_Microscope>
  </Instrument_Setup>
  <Datasets>
    <Subgroup>
      <Name>kazue-1tiff</Name>
      <Description>Block 2-7-93, hemangioma-20, slide 37</Description>
      <Path>kazue-1tiff</Path>
      <Image_Stack Created_Date="1993-06-28" ID="000001-1" URN="urn-
  3:hms.count:20030528-85983" Thumbnail_HREF="http://nrs.harvard.edu/urn-
  3:hms.count:20030528-85995" Fullsize_URN="http://nrs.harvard.edu/urn-
  3:hms.count:20030528-85989" Image_Type="greyscale" Planar_Configuration="1">
        <Path>kazue-1tiff/1</Path>
        <Creator>Rogers</Creator>
        <SiteRef ID="6" Name="HSPH"></SiteRef>
        <Subject>Adolescence</Subject>
        <Subject>Child</Subject>
        <Subject>Child, Preschool</Subject>
        <Subject>Collagenases</Subject>
        <Subject>Endothelial Growth Factors</Subject>

```

```

<Subject>Fibroblast Growth Factor</Subject>
<Subject>Gelatinase B</Subject>
<Subject>Glycoproteins</Subject>
<Subject>Hemangioma</Subject>
<Subject>Human</Subject>
<Subject>Infant</Subject>
<Subject>Infant, Newborn</Subject>
<Subject>Lymphokines</Subject>
<Subject>Mast Cells</Subject>
<Subject>Neoplasm Staging</Subject>
<Subject>Nuclear Proteins</Subject>
<Subject>Proliferating Cell Nuclear Antigen</Subject>
<Subject>Tissue Inhibitor of Metalloproteinases</Subject>
<Subject>Tumor Markers, Biological</Subject>
<Caption>Low mag microvessels in biopsy tissue</Caption>
<Rights>President and Fellows of Harvard University</Rights>
<Series Thumbnail_TIFF="ch01f00020.tif">

  <Z_Series Z_Step="1.0 micron" Number_Of_Steps="130">

    <File_Naming_Convention>n+1</File_Naming_Convention>
  </Z_Series>
  <T_Series T_Step="None" Number_Of_Steps="0">

    <File_Naming_Convention>Not applicable</File_Naming_Convention>
  </T_Series>
  <XY_Series X_Step="None" Y_Step="None" Number_Of_Steps="0">

    <File_Naming_Convention>Not applicable</File_Naming_Convention>
  </XY_Series>
  <Channel_Series Channel_List="None" Channel_Unit="Not applicable">

    <File_Naming_Convention>Not applicable</File_Naming_Convention>
  </Channel_Series>
</Series>
<PreparationRef ID="1" Name="Single channel 488"></PreparationRef>
<InstrumentRef ID="1" Name="Confocal"></InstrumentRef>
<Sampling>purposeful search</Sampling>
<Viewing_Note>fluorescence</Viewing_Note>
<Artifact>none noted</Artifact>
</Image_Stack>
</Subgroup>
</Datasets>
</Project>

```