

CIDC

CORNELL INSTITUTE FOR DIGITAL COLLECTIONS

First Annual Report

1 July 1997 - 30 June 1998

I. Cornell Institute for Digital Collections: Mission and Program Goals

The Cornell Institute for Digital Collections (CIDC) was formally established on September 1, 1998. Its mission is to explore the use of emerging technologies to expand access to cultural and scientific sources and to support the use of these resources on campus and globally through the development and management of distinctive digital collections. This mission is being realized through academic, corporate, and public partnerships, national and international. In this critical area of educational and public concern, CIDC is contributing a unique combination of curatorial, instructional, research, and managerial expertise.

Current program goals for the Institute are:

1. Create and maintain digital resources representing the content of Cornell University collections, in combination with visual, audio, and textual sources from around the world.
2. Foster the incorporation of these new resources into the classroom, laboratory, studio, and home.
3. Conduct applied research into the use and utility of digital technology to archive and disseminate important cultural and scientific information.

4. Educate and train in the employment of digital technologies for creating, publishing, and using electronic resources.
5. Investigate new models for administering access and use of digital collections.
6. Cooperate internationally with other universities, corporations, foundations, countries, and standards organizations in the development and administration of digital resources.

II. Executive Summary

Confirmation of expanded funding was received in August 1998. Until that point, staffing had been maintained at a minimal level. Initial attention was devoted to extending appointments of some existing staff, hiring new staff, and establishing a management structure necessary for administering an expanded program. Additional Library space was assigned, and a more robust technical infrastructure was implemented. This expansion was completed in three months.

Continuing on the model of its predecessor, the Digital Access Coalition, the Cornell Institute for Digital Collections has focused during its first year on the testing and refining of technologies for creating digital surrogates of cultural images and artifacts. The Digital Access Coalition had worked directly with teaching faculty in testing new imaging technologies and developing effective methods for classroom use. The Coalition served as communal sounding board for teachers, curators, and technologists exploring these new possibilities. The Coalition brought several of their initial visions to reality, validating the promise of this new approach.

While maintaining existing commitments, the Cornell Institute for Digital Collections expanded its partnership with the Herbert F. Johnson Museum of Art. CIDC joined with the Johnson in an ambitious effort to create high-resolution copies of over 85% of the Museum's holdings and to provide networked access to those images and accompanying information. Due to its initial success in generating high-quality images in a production environment, the Museum Online Project is beginning to acquire "Mecca" status as museum professionals travel to Cornell to learn from our experience.

Although the Museum Online Project and associated activities have occupied much of our attention, other new enterprises have developed, and CIDC has firmly established its role as a facile and creative partner in campus-wide, national, and international initiatives. While we will continue our involvement in these areas, we propose to concentrate more on the means for disseminating access to digital resources and managing the distribution process.

Principal attention will be devoted to the development of effective distribution and management models and administering access to digital products. Areas of concentration will include: licensing, intellectual property, product development models, distribution models, branding, marketing, and e-commerce. If broadly endorsed, we would seek to develop a high-quality, educationally-focused selection of materials, employing rich content, outstanding technical capabilities, and excellent design. We will explore the development of a diverse range of partnerships, both in the development and inclusion of sources and in the distribution and marketing of resources. The overall goal is to establish Cornell, through the CIDC, as a principal non-profit Internet "portal" for high-quality educational productions.

Total expenditures at the end of Year 1 (\$554,996) equaled only 80% of original projections. This reflects the initial delay in program expansion and will result in the certain expenses scheduled for 1997-1998 being deferred to 1998-1999. Incoming revenues for Year 1 totaled \$810,305, leaving a balance of \$255,309 on June 30, 1998. Projected expenditures for 1998-1999 are \$688,516, a 75% increase over original Year 2 projections. This increase reflects deferred expenditures and projected increases to support new initiatives. While revenues for Year 2 will exceed current projections if one or more federal funding proposals is successful, presently a \$57,982 deficit is projected. This would leave at the end of Year 2 a remaining balance of \$197,327.

III. Organization and Administrative Developments

A portion of the first year was spent organizing CIDC and managing its transition from the Digital Access Coalition. The University Library has provided vital support for the direction, administration, and space needs of the Institute. This includes the salary of the director, the support of the Library's financial and human resource management systems and personnel, and office space in the Carl A. Kroch Library (c. 500 sq. ft., plus regular use of meeting and conference areas). Additional space for CIDC activities was provided by the Interactive Media Group in the Department of Communications and the Herbert F. Johnson Museum of Art.

1. Staffing

The success of CIDC is to a large part dependent on its ability to gather and utilize curatorial, instructional, research, and managerial expertise. The first task facing CIDC, therefore, was to gather a talented staff. In some cases, specialists who had worked on the grant-funded projects of the Digital Access Coalition were available for hire; in other cases new staff had to be found. Short descriptions of primary staff involved with the Institute follow:

H. Thomas Hickerson, the Associate University Librarian for Special Collections, serves as Director of the Institute in addition to his other regular duties. This year, he also served six months as Acting Associate University Librarian for Library Information Technology. Hickerson organized and served as co-director of the Digital Access Coalition. He was a 1993 recipient of an Andrew W. Mellon Research Fellowship to investigate digital access to research collections on college and university campuses and has spoken and written extensively on library and archival automation. He was named a Fellow of the Society of American Archivists in 1989 and is presently incoming Vice President/President-Elect of the Society. His publications this past year include: *Developing Collaborative Structures for Expanding the Use of University Collections in Teaching and Research: A Case Study and Teaching Notes* (Co-authored with Elaine D. Engst. Chicago: Society of American Archivists, 1998); "Exploring Models for Collaborative Development and Management of Digital Collections" (*To Scan Or Not To Scan: What Are the Questions, Proceedings of a SOLINET Conference on Digitizing Projects for Libraries and Archives*. Atlanta: SOLINET, 1997); and "Realizing New Means: Networked Access to Research Collections" (*Museums and the Web 97: Selected Papers*. Pittsburgh: Archives & Museum Informatics,

1997). Hickerson provides general direction to CIDC programs.

Peter B. Hirtle was appointed Assistant Director of CIDC after acting as the Manager of the Digital Access Coalition. Prior to his arrival at Cornell, Hirtle worked at the National Archives and Records Administration, first for the Technology Research Staff (where he helped complete their last digital imaging report), and then as coordinator of electronic public access for the agency. He is a member of the governing Council of the Society of American Archivists and of the Society's Executive Committee. He is a frequent speaker at professional meetings, and during the past year, he conducted or participated in five workshops providing basic and/or advanced training in the use of the digital imaging technologies. He participated in the preparation of six funding proposals seeking support from federal, corporate, and foundation programs, and involving various collaboration among Cornell's Library, Department of Computer Science, Interactive Media Group, and Johnson Museum, and also the University of California at Berkeley, Harvard University, and the Museum Digital Licensing Cooperative. He is co-manager of the Museum Online Project, manages the IMAG/CU project, and serves in a consulting role to Cornell's Archival Electronic Records Study, the Digital Library Federation's Making of America II Project, and the SagaNet Project, all based in the Library's Division of Rare and Manuscript Collections. Hirtle is also responsible for the daily operations of CIDC.

Oliver B. Habicht was hired in December 1997 as CIDC Systems Engineer. Prior to his appointment, Habicht had spent seven years as a technical advisor to Cornell Information Technologies. He has a strong background in Internet multi-media technologies and production client-server systems. His responsibilities involve creating and maintaining the technical environment that supports CIDC's many activities and coordinating software implementations and applications development.

Robert Rieger, the coordinator of the Interactive Media Group, also serves as CIDC's Educational Projects Manager. He has assisted CIDC with proposal preparation, research design, data collection, analysis, evaluation, and reporting. He also assists with teaching and planning. Rieger has played the lead role in the Plantations Project.

Noni Korf Vidal is Interface Design Manager. She has been the principal designer of the Web interfaces for many of the databases and online exhibits previously implemented by the Digital Access Coalition and the Division of Rare and Manuscript Collections. [The online version of the exhibition, "Paper, Leather, Clay, and Stone: The Written Word Materialized," (<http://rmc.library.cornell.edu/Paper-exhibit/default.html>) was named a Biblio Top Ten Web Site in June 1998.] She has also played an active role in promoting the use of these databases by faculty and students. She coordinates CIDC participation in the Death of the Father Project and traveled to Bellagio, Italy to take part in a meeting of international participants.

Rebecca W. Davidson, Visual Collections Cataloger, completed cataloging for images in the Utopia database for the study of the art and Architecture of the Italian Renaissance. She combines a decade's experience with library cataloging and information systems with a doctorate in the History of Architecture and Urbanism. She is the chief cataloger for the Museum Online Project.

2. Technical Infrastructure

CIDC's information technologies consist of server and client computers connected by an Ethernet Local Area Network (LAN) to Cornell's Internet fiber-optic backbone. Three servers are at the heart of the CIDC infrastructure. One server, a Pentium Pro 220 Mhz machine with 64 MB of RAM, hosts the web and FTP sites. A second identical server is used to back-up the first server, and for testing. The third server runs an Oracle application managing records of the museum holdings and several slide and image libraries at Cornell. It has four 200 Mhz Pentium II processors, 256 MB of RAM, and 45 GB of disk storage. A Windows NT workstation with an oversized monitor is used to drive a high-resolution flatbed color scanner and CD-ROM recorder. Two other NT workstations, both with 300 Mhz Pentium II processors, are used by CIDC staff for daily production work. A black-and-white laser printer and a color inkjet printer are connected to the LAN as well. CIDC is scheduled to receive two additional servers and workstations in 1998/9 as part of a gift from the Intel Corporation.

In addition to this equipment in the central CIDC offices, additional equipment has been acquired for the Museum Online Project. A PowerPhase digital camera on a ZBE Satellite camera stand is used to capture images. The camera is driven by a 300 Mhz Macintosh 9600 with 8 GB of storage and supplementary Apex storage cartridges. Images are edited and maintained on a 100BaseT Ethernet LAN consisting of a server and two workstations. All machines are 233 Mhz Pentium II machines running either Windows NT server or workstation. A dual drive CD recorder is also used.

3. Publicity

A CIDC Web site (<http://cidc.library.cornell.edu>) came online soon after CIDC was established. The design of a CIDC prospectus was also completed. [Copies included.]

4. Advisory Board

The first meeting of the Advisory Board will be held in October 1998.

IV. CIDC Projects

Museum Online (program goals 1, 2)

With the substantial support of an anonymous donor, CIDC, in conjunction with the Herbert F. Johnson Museum of Art, began in January 1998, an ambitious project to prepare digital surrogates for 85% of the items in the Museum. A digital photo studio was selected and installed, and two digital photographers, a part-time systems support specialist, and two catalogers were hired to capture the images, prepare derivatives suitable for delivery on the Web, and provide access points that will make searching for images possible. As of July 1, 1998, over 4,000 items (15% of goal) from the Museum's holdings were available in digital form. By putting the collection online, students, faculty, visitors, and users on the Internet will have access to the riches of one of the country's major university museums, only a fraction of which can be exhibited at any one time. Once converted and accessible, the digital collection can be integrated into Cornell's curriculum, and planning is underway on how the museum images could be used in local public schools.

IMAG/CU (program goals 1, 2)

CIDC has agreed to develop and maintain a database for a shared image collection compiled from slides from the slide libraries of the College of Architecture, Art, and Planning and the Department of History of Art; digital surrogates from the Herbert F. Johnson Museum of Art; and photographs from the Division of Rare and Manuscript Collections in the Cornell University Library. Searches conducted across all four collections will make it a valuable interdisciplinary teaching and research tool. In support of this effort, CIDC has installed museum management software that will be used by all the participants, and has overseen the conversion of data from the four repositories. CIDC has also contracted for delivery of a Web interface to the shared database, to be delivered in September.

Plantations Project (program goal 2)

Robert Rieger and other staff of the Interactive Media Group (IMG) have developed multiple prototypes designed to showcase the use of mobile computers in learning. One project undertaken under the auspices of the CIDC tests whether highly portable, multimedia computing resources improve learning experiences. IMG and CIDC worked in conjunction with the Plantations to create a series of digital resources, including hand-held computerized inventories of the Plantations and a web site for electronic visitors. In part as a consequence of this initiative, a private donor has provided joint funding for the Plantations and the New York Botanical Gardens to develop networked plant databases.

Museum Digital Licensing Collective (MDLC) (program goals 4, 5, 6)

CIDC has joined this collaborative initiative, along with the American Association of Museums, Sun Microsystems, and the University of California at Berkeley, to develop a nationwide collection of digital images of museum holdings that could operate both in a

not-for-profit as well as a for-profit manner. The Cornell University Librarian was appointed to its governing board. Current plans call for the CIDC to take the lead in the development and implementation of imaging standards. In addition, if funding is secured, the CIDC will provide training through workshops, manuals, and online tutorials.

Death of the Father (program goals 2, 6)

What happens to a culture after the death of a totalitarian leader such as Stalin, Hitler, or Mussolini? This international anthropological project explores the closure of political authority following the death of these patriarchal figures. CIDC has contributed to the development of networked multimedia infrastructure that permits written text to be integrated with digital images and film for a fuller, more penetrating examination of the topic. The project will test the viability of interdisciplinary collaboration and classroom discussion on a global scale and breaks new ground in creating multimedia course materials on the World Wide Web.

Making of America II (program goals 1, 3, 6)

The Making of America II (MOA2) is a project sponsored by the Digital Library Federation and funded by a grant from the National Endowment for the Humanities. Its purpose is to extend research and demonstration projects that have begun to develop best practices for the encoding of intellectual, structural, and administrative data about primary resources housed in research libraries. CIDC is representing the Cornell University Library on the MOA2 development team. During this past year, CIDC has advised on the nature and structure of primary source documents. This next year will see the construction of a testbed of data and images.

Maintenance of Digital Access Coalition Projects (program goals 1, 2, 5)

CIDC absorbed the maintenance of pilot projects undertaken by its predecessor, the Digital Access Coalition. Among the projects maintained by CIDC are:

UTOPIA, an online database of over 5,000 images of European Renaissance art objects, artifacts, architecture, and gardens;

The Museum Educational Site Licensing Project (MESL), a collaborative effort of the Getty Information Institute, seven museums, and seven universities. MESL provides access to over 9,000 images of paintings, photos, textiles, ceremonial objects, and other cultural artifacts;

The Louis Agassiz Fuertes Collection, a collection of 2,600 images of artwork and personal papers of notable ornithological painter, Louis Agassiz Fuertes. The site tests the integration of text, images, and other media in one common environment.

V. Funding Initiatives

The funding received to establish the Institute has made it possible to begin the projects listed above. The expansion of activities will require additional funding, however. As a consequence, the CIDC has spent significant effort in exploring ways to supplement existing resources. Funding would assist in supporting existing efforts campus-wide or allow the adoption of new initiatives. Additional proposals are currently being planned for submittal in 1998 to accelerate the implementation of new programmatic directions by CIDC.

CIDC staff were active in preparing the following proposals in 1997/1998; the Intel proposal has already been successful.

- Intel Corporation: "Technology for Education 2000" grant. The Museum Online Project was as one of the component parts in Cornell University's successful "Technology for Education 2000" grant program. CIDC has received from the grant almost \$35,000 in hardware and software, with almost \$20,000 in additional hardware going to our Museum collaborators. Additional equipment valued at \$58,000 will be received in 1998/1999.

- Institute of Museum and Library Services (IMLS), "Digital Imaging of Three-Dimensional Objects for Museums and Libraries," \$251,206 request.

The request to IMLS would pay for completion of the Museum Online project. Included are funds to research the current state of the art in the capture of three-dimensional objects, purchase an appropriate camera, and capture digitally artifacts in the Museum

- Institute of Museum and Library Services, "Museum Digital Licensing Collective Testbed Project," \$497,375 request.

This grant application, crafted in conjunction with the Peabody Museum of Archaeology and Ethnology at Harvard University, the University of California at Berkeley, and the Museum Digital Licensing Collective, would fund the first year of an MDLC testbed. Cornell's subcontract would include funds for developing best practices for digital imaging; preparing workshops on digital imaging for museums; and developing an evaluation of the MDLC testbed.

- Telecommunications and Information Infrastructure Assistance Program, "OMNI Online" CIDC brought together the Interactive Media Group, the Herbert F. Johnson Museum, and the Tompkins

County BOCES school district. Together they developed a grant application to develop and test innovative methods of utilizing digitized collections in K-12 education by providing an online, interactive learning environment to rural, technologically under-served audiences.

· Council on Library and Information Resources, "Risk Management of Digital Information," \$75,758 request.

CIDC staff served as technical consultants in the preparation of this grant application submitted by the Cornell University Library. The grant seeks to identify the preservation risks associated with two digital formats, develop a risk assessment tool, and investigate practical preservation procedures to implement a format migration strategy.

· National Science Foundation, "Digital Libraries II Initiative," CIDC staff contributed to the preparation of a grant application to the National Science Foundation for funding under its Digital Libraries program. The proposed research, to be carried out in the Department of Computer Science, the Cornell University Library, and the Interactive Media Group, would develop the specifications for the infrastructure of a distributed digital library. The specifications would take into account the needs of libraries to preserve digital information over time, \$3,714,991 request.

VI. Continuing Commitments and Future Development, 1998/1999

1. On-Going Projects (as of July 1, 1998):

· MUSEUM ONLINE PROJECT

Conversion and Cataloging

This Project will continue for eighteen more months, at which time 85%+ of the Museum's holdings and descriptive information and indexing will be accessible in digital form. At that time, the equipment that has been acquired will be available for other CIDC initiatives.

Access and Use

The Museum database will be made available for classroom and research use through the IMAG/CU project described below.

· IMAG/CU PROJECT

CIDC is committed to maintaining the central database for the IMAG/CU project for at least four years (1998-2001). In 1998/99, we will develop a Web interface for this system to facilitate direct access by users. CIDC is coordinating the use of common data standards and will assist in local implementations by the participating units. CIDC will also explore the development of selected thematic products derived

from the central database. Other repositories on campus will be invited to employ the system.

'' **Additional Projects**

Existing commitments to development of the Plantations database and Web site, "The Death of the Father" Project, Making of America II, and SagaNet, will be maintained.

2. Proposed New Initiatives

New initiatives for 1998-1999 can be categorized under the concept of INTERNET PUBLISHING. In the past year we have focused considerable attention on technical requirements for high-resolution capture in a production environment. We have continued our ongoing concern for student use of cultural and scientific information in digital form, while expanding beyond just the college classroom. We have disseminated our findings through presentation and publication, through workshop instruction, and by involving Cornell students in developing interactive media. While we will continue our involvement in these areas, we propose to focus on the best means for disseminating access to digital resources and managing the distribution process.

'' **INTERNET PUBLISHING**

Principal attention will be devoted to the development of effective distribution and management models and to administering access to digital products. Areas of focus will include: licensing, intellectual property, product development models, distribution models, branding, marketing, and e-commerce. If broadly endorsed, we would seek to develop a high-quality, educationally-focused selection of materials, employing rich content, outstanding technical capabilities, and excellent design. We would explore the development of a diverse range of

partnerships, both in the development and inclusion of sources and in the distribution and marketing of resources. We might draw from private as well as institutional collections and work in collaboration with both educational and corporate partners. We would investigate the distribution of our products through non-profit consortia like MDLC, private companies like Primary Source Media, or through Cornell-specific channels. Many of these resources could be distributed freely; others would be sold through purchase or subscription; and others might be used as components in distance education programs. The overall goal would be to establish Cornell, through the CIDC, as the chief non-profit Internet "portal" for high-quality educational productions. Accompanying technical research should focus on technical platforms, derivative production, compression schemes, and networking models. We would also explore the use of various encoding schemes and display technologies. All of these investigations would address the need to incorporate sound, motion, text, and still images into integrated resources.

We cannot address all of these possibilities at once. After review of this concept by the Advisory Board, we will commission one or more consultants to examine the most promising avenues for initial pursuit and to suggest possible business models. We would hope to have initiated these new activities by December 1, 1998.

VII. Budget

As reflected in the accompanying statements of Income and Expenditures and Budget Projections, total expenditures at the end of Year 1 (\$554,996) equaled only 80% of original projections. This reflects the initial delay in program expansion and will result in certain expenses scheduled for 1997-1998 being deferred to 1998-1999. Incoming revenues for Year 1 totaled \$810,305, leaving a balance of \$255,309 on June 30, 1998. Projected expenditures for 1998-1999 are \$688,516, a 75% increase over original Year 2 projections. This increase reflects deferred expenditures and projected increases to support new initiatives. While revenues for Year 2 could exceed current projections if one or more federal funding proposals is successful, presently a \$57,982 deficit is projected. This would leave at the end of Year 2 a remaining balance of \$197,327