

# Digital Dictionaries of South Asia

Funded by the U.S Department of Education under Title VI, Section 605,  
October 1999 through September 2002

---

**PROGRAM**

International Research and Studies Program,  
International Education and Graduate  
Programs Service, U.S. Department of  
Education, CFDA No. 84.017A

**FUNDING REQUESTED**

\$453,071 for three years

**PROJECT DATES**

Oct. 1, 1999 - Sept. 30, 2002

---

Project Director's Signature

**TITLE OF PROJECT**

Digital Dictionaries of South Asia

**APPLICANT**

The University of Chicago  
970 East 58<sup>th</sup> Street  
Chicago, Illinois 60637  
773-702-8602

**PROJECT DIRECTOR**

James H. Nye  
Director, South Asia Language  
and Area Center  
5848 University Avenue, Kelly 313  
Chicago, Illinois 60637  
773-702-8430

---

**SUMMARY OF PROPOSED PROJECT**

For language learning and instruction, few resources are more crucial than dictionaries. This project aims to make high-quality dictionaries in each of the twenty-six modern literary languages of South Asia universally available in digital formats. At least thirty-two dictionaries will be converted from printed books, often multi-volume, to electronic resources. A wide variety of users will benefit from access to electronic dictionaries via global media such as the World Wide Web. Not only the academics whose study of Indic languages has long been supported by the Department of Education, but also American-born learners of South Asian heritage, and individuals around the world will profit.

A well-developed plan and the considerable experience of key personnel ensure that the project's objectives will be met. The Project Director and two Co-Directors have been at the forefront of recent initiatives to improve global access to South Asian materials through deployment of current technologies. Their direction is supported by an infrastructure of institutional and technological resources. Of particular importance is the expertise in digitization to international standards already achieved by the University of Chicago's American and French Research on the Treasury of the French Language (ARTFL). The quality of the project's academic direction is ensured by an Advisory Panel of eminent scholars in the field of South Asian linguistics.

The project will be implemented in stages, subject to ongoing and final evaluation. Each dictionary will be selected according to merit, converted to digital format, reviewed for accuracy by South Asian scholars, implemented on the World Wide Web, and tested for trouble-free operation. The double-keying technique of conversion is notable for its efficiency, accuracy, and cost-effectiveness. Conferences for critique and dissemination of project information among scholarly and professional bodies insure ongoing review of the program's activities.

The Digital Dictionaries of South Asia project is one element in a larger cooperative movement to maximize access to cultural records from the Indian subcontinent. In its final format, it will make vernacular dictionaries available in three versions: readable on the World Wide Web, downloadable by file transfer protocol (ftp), duplicated at cost in the form of compact discs (CDs). The project offers unprecedented and unparalleled promise of reliable long-term access to accurate dictionaries. It will remedy the current lack of reliable and scholarly on-line lexical resources for South Asian studies.

## Table of Contents

|           |   |    |
|-----------|---|----|
| Part I:   | Application for Federal Assistance (Standard Form 424)                            |    |
|           | Abstract  |    |
| Part II:  | Budget Information – Non-Construction Programs (ED Form 524)                      |    |
|           | Itemized Budget Breakdown   |    |
|           | Budget Documentation  |    |
| Part III: | Application Narrative   |    |
|           | Introduction .....  | 1  |
|           | 1. Need for the Project.....  | 1  |
|           | 2. Usefulness of the Product to Others .....                                      | 5  |
|           | 3. Expected Contribution to Mandate under Section 605 .....                       | 11 |
|           | 4. Account of Related Materials .....   | 11 |
|           | 5. Likelihood of Achieving Results .....  | 12 |
|           | 6. Quality of Personnel .....   | 15 |
|           | 7. Plan of Operation.....   | 17 |
|           | 8. Budget and Cost Effectiveness .....  | 23 |
|           | 9. Plan of Evaluation.....  | 23 |
|           | 10. Adequacy of Resources .....   | 25 |
|           | 11. Description of Final Format.....  | 25 |
|           | 12. Provisions for Pretesting.....  | 26 |
| Part IV:  | Appendixes  |    |
|           | 1. Institutional Participants .....   | 28 |
|           | a) University of Chicago.....   | 28 |
|           | b) The ARTFL Project at the University of Chicago.....                            | 29 |
|           | c) Columbia University.....   | 30 |
|           | d) Triangle South Asia Consortium .....   | 31 |
|           | 2. Curricula Vitae for Key Staff .....  | 32 |
|           | a) James H. Nye.....  | 32 |
|           | b) David S. Magier .....  | 34 |
|           | c) Tony K. Stewart.....   | 36 |
|           | 3. Letters of Commitment.....   | 38 |
|           | 4. Dictionaries to be Converted .....   | 41 |
|           | 5. Specimen ARTFL Dictionary Web Site .....                                       | 45 |
|           | 6. Statements on Digital Standards.....   | 46 |
|           | a) The Unicode Standard, Version 2.0.....   | 46 |
|           | b) The Dublin Core: A Simple Content Description Model for Electronic Resources.. | 47 |
|           | c) Standard Generalized Markup Languageand the Text Encoding Initiative.....      | 48 |
|           | 7. ICOLC "Guidelines for Statistical Measures of Usage" .....                     | 49 |
|           | 8. Impact of Technology on South Asian Studies .....                              | 51 |
|           | 9. Data Entry by Double Keying .....  | 52 |
|           | 10. Response to GEPA Section 427 .....  | 53 |
|           | Assurances  |    |
|           | Non-Construction Programs (Standard Form 424B)                                    |    |
|           | Certifications Regarding Lobbying; Debarment, Suspension and Other                |    |
|           | Responsibility Matters; and Drug-Free Workplace Requirements (ED80-0013)          |    |
|           | Certification Regarding Debarment, Suspension, Ineligibility and                  |    |
|           | Voluntary Exclusion – Lower Tier Covered Transactions (ED80-0014)                 |    |

## **Introduction**

The University of Chicago, Columbia University, and the Triangle South Asia Consortium in North Carolina propose a three-year project to create and disseminate electronic dictionaries. A panel of language experts will identify key, established dictionaries which currently exist in print for each of the twenty-six modern literary languages of South Asia. At least one multilingual dictionary will be selected for each South Asian language; in the case of more frequently taught languages, an additional monolingual dictionary will also be chosen. After the best extant resources have been identified, a team of project staff will then convert the dictionaries to digital formats. The results of this conversion will be available to readers by means of a searchable site on the World Wide Web, by means of standard file transfer protocol, or by compact disc. There will be no charge for access via the Internet and the compact discs will be available for the cost of duplication and mailing. This project will make a major contribution to the larger international lexical infrastructure while providing high quality resources for use by scholars and lay readers throughout the world.

### **1. Need for the Project**

#### **Need for Dictionaries**

This project will build a vital part of the **digital architecture for language learning and instruction**. Few resources are as important for language knowledge as good dictionaries. For academic users, the ability to consider not simply the most commonly understood meaning of a word or phrase but its diverse social connotations is of great importance. Without the benefit of high quality dictionaries the critical nuances of great literature and detailed scholarship are difficult to express or comprehend, for native speakers as well as other readers. Currently in the U.S. there are at least eighteen South Asian literary languages being taught in universities. In addition, there are a significant number of second-generation South Asians living in the U.S. interested in maintaining or improving upon their linguistic heritage. This potential pool of users is increasing considerably so that according to the 1990 census one in every 270

Americans describes themselves as having ancestors from South Asia. The need for dictionaries is also increasing in part because of the efforts of the Association for Asian Studies' South Asia Council to promote remote language instruction for the least-commonly taught languages as well as for students and scholars living in isolated areas. Outside of the U.S. South Asia currently comprises more than one-fifth of the world's population and the use of Internet is expanding rapidly. In order to meet this growing need for dictionary resources, a few well-intentioned enthusiasts have already mounted a small number of badly prepared dictionaries on the Internet. These inadequate and sometimes transitory efforts are clearly insufficient but reflect the growing need for such resources.

#### Native Speakers of Languages Addressed by this Project<sup>1</sup>

| Language  | No. of Speakers |
|-----------|-----------------|
| Assamese  | 13,000,000      |
| Baluchi   | 3,600,000       |
| Bengali   | 176,000,000     |
| Divehi    | 220,000         |
| Dogri     | 2,000,000       |
| English   | 11,119,000      |
| Gujarati  | 40,200,000      |
| Hindi     | 341,000,000     |
| Kannada   | 32,000,000      |
| Kashmiri  | 4,100,000       |
| Konkani   | 1,800,000       |
| Maithili  | (Hindi dialect) |
| Malayalam | 29,500,000      |
| Manipuri  | 1,252,000       |
| Marathi   | 62,000,000      |
| Nepali    | 11,700,000      |

| Language     | No. of Speakers           |
|--------------|---------------------------|
| Newari       | 690,000                   |
| Oriya        | 27,000,000                |
| Pali         | (classical <sup>1</sup> ) |
| Panjabi      | 94,000,000                |
| Pashto       | 15,000,000                |
| Persian      | 1,000,000                 |
| Prakrit      | (classical <sup>1</sup> ) |
| Rajasthani   | (Hindi dialect)           |
| Sanskrit     | (classical <sup>1</sup> ) |
| Sindhi       | 16,500,000                |
| Sinhala      | 12,800,000                |
| Tamil        | 63,000,000                |
| Telugu       | 68,000,000                |
| Urdu         | 54,000,000                |
| <b>TOTAL</b> | <b>1,080,461,000</b>      |

More than 86% of the total South Asian population has one of these languages as their mother tongue.

In addition to direct use by scholars and students these dictionaries will be an important component of future efforts to produce full-text files of language resources. The dictionaries will allow for the testing

<sup>1</sup>Combined populations in South Asia are given for languages spoken in more than one country. Statistics are based on the 1991 censuses for Bangladesh, India, the Maldives, Nepal, Pakistan, and Sri Lanka.

Federal funds will not be used to create digital dictionaries for the classical languages Pali, Prakrit, and Sanskrit. Instead matching funds from Columbia University are already committed for a Sanskrit dictionary. The Project Director will seek other support during this three-year project for creation of Pali and Prakrit dictionaries.

and verification of such files in order to ensure their accuracy. Existing dictionaries are clearly insufficient (detailed examples of the deficiencies are described in section 4 below) for this purpose as well. Commercial firms are unlikely to provide for this need. For example, Oxford University Press considers the commercial potential for an electronic version of their publication, *The Oxford Hindi-English Dictionary* by R. Stuart McGregor, too small to warrant their effort, even though the text of the dictionary is available in machine-readable form.

### **Significance of South Asian Studies**

South Asian studies encompasses the subcontinent south of the Himalayan mountain range and comprises India, Pakistan, Bangladesh, Nepal, Sri Lanka, Bhutan, and the Maldives Islands. The site of major civilizations for the past four millennia, the intellectual domain is enormous both in its geographical area and chronological duration. The size of the subject is matched by the American contribution to the study of this region. Since 1960 scholars from North America have done more to enhance knowledge of South Asia than their colleagues from any other region save the subcontinent itself. This assessment is based on the number and quality of books and articles written, Ph.D.s awarded, courses taught, and language instruction offered.<sup>2</sup> This growth of knowledge has been supported by forty-one years of funding from the U.S. Department of Education for foreign language and area training. This sustained support has produced an intelligent and demanding body of scholars and policy makers with demonstrable capacities to use the languages of South Asia and comprehend the nuances of South Asian civilizations. They expect and deserve more extensive and flexible access to electronic resources in order to train succeeding generations of researchers and leaders as well as to further their own inquiries.<sup>3</sup> This project will address that need.

South Asian studies in the United States has historical roots in nineteenth-century philological and religious studies focused on classical India. In the aftermath of World War II there was a marked change.

---

<sup>2</sup> N. Gerald Barrier, *India and America: American Publishing on India, 1930-1985* (New Delhi: Manohar; American Institute of Indian Studies, 1986).

<sup>3</sup> Please see Appendix 8 for a succinct statement on the inadequacy of digital resources for South Asian studies, extracted from the 1996 publication *Scholarship, Research Libraries, and Global Publishing*.

The war highlighted America's ignorance about the civilizations of South Asia (as well as other world areas.) To remedy this shortcoming systematic efforts were made to develop programs of study, first at a few universities and then at many more by the 1960s. These new South Asia programs had distinctive intellectual characteristics and disciplinary orientations shaped by founding faculty. Yet, in general, there was a strong emphasis on the social sciences and on regional rather than classical languages.

The **social sciences** have remained vital and continue to attract the interest of new students and of scholars without backgrounds in South Asian studies. Some researchers have been engaged by the writings of South Asian scholars addressing issues such as colonialism. They have viewed the sources from varied perspectives, utilizing resources in the regional languages of the subcontinent. Others have been interested in more novel approaches to understanding such as those centered on popular culture (or public culture), cinema, and gender studies. During this period of economic liberalization and increased foreign investment, economics and political science continue to attract the attention of many. The **humanities** are also vital areas of research related to South Asia. As an example, scholars of regional literatures are writing new histories of the enormous and varied literary traditions in the subcontinent, challenging previous presuppositions of literary and language boundaries, chronological divisions, nationalism, and canon that undergirded earlier publications.

While most researchers come from a discipline – both as a departmental home and a methodological disposition – there is increasing interest in **interdisciplinary studies and trans-regional research** on South Asia. One can point to investigations about the environment, nationalism, human rights, and migration as examples of topics commanding contemporary and future attention. In addition to their wide geographical scope, these inquiries often have a global impact. To study South Asia through the social sciences, humanities, or interdisciplinary studies, it is critical to have a solid knowledge of the languages of the subcontinent.

The need for this project on South Asia is acutely felt in **research libraries**. Some of the best dictionaries are out of print but remain in high demand. One good example is the fourth edition of *J. P. Fabricius's Tamil and English Dictionary*. This revised and enlarged edition, based upon the first edition

of 1770, is highly valued by Tamil students and teachers. Other dictionaries are infrequently held in the U.S. because the languages are among the least commonly taught. For less commonly consulted publications inter-library loan remains an acceptable alternative to widespread duplication in acquisition. However, for core reference resources such as dictionaries, many of which do not circulate, widespread demand requires alternative solutions that can provide a number of readers with important works in a timely fashion. The direct delivery of the highest quality dictionaries to readers via the Internet as proposed in this project is such an alternative.

**Beyond the primary audiences** of scholars at South Asia National Resource Centers and other higher education institutions, other important bodies of readers will be well served by the electronic dictionaries prepared and disseminated under this project. These include:

- school teachers;
- cultural institutions, especially museums and libraries;
- governmental and non-governmental organizations (NGOs);
- businesses; and
- news reporters and others in the media.

Through this project, the needs of a **wide array of U.S. citizens** for ready access to lexical resources will be met. (Please see Section 2 for examples of how the resources might be used.) They will be able to turn to the project's site on the World Wide Web for free and dependable access to dictionaries of the highest accuracy with powerful, straightforward interfaces.

## **2. Usefulness of the Product to Others**

This project's **careful design** ensures that the resulting electronic resources will be useful to the targeted audiences. An Advisory Panel of distinguished linguists and language teachers will select the best printed dictionaries for consideration. They will use selection criteria specifically related to the value of the lexicons for teaching and research. Once selected, paper copies of the dictionaries will be delivered to a subcontractor responsible for keying the text into electronic form and adding necessary tags to make the files useable as structured databases. Keying and tagging using Standard Generalized Markup Language (SGML) will be done using standard double keying techniques to ensure extremely high levels

of accuracy.<sup>4</sup> After each electronic dictionary file is delivered, project staff and an independent contractor in India will test the files to make certain that the contractually specified levels of accuracy are maintained and that application of SGML tags for data elements is correct. An expert team of programmers at the University of Chicago will mount the dictionaries on a World Wide Web site with a powerful and yet intuitively simple interface for use by readers. Further, the lexical data will be made available to readers by downloading from the Web site or on compact disc, if they prefer. As noted in Section 7, the project will ensure that copyright laws are carefully followed in the conversion of printed dictionaries to electronic form. Finally, careful and thorough evaluation of the digital dictionaries and the means of their dissemination will allow project staff to fine tune the methods of delivering resources to readers, if that is necessary.

This project is significant because it **effectively addresses needs** registered by national foundations and associations of scholars and librarians. Funded by The Andrew W. Mellon Foundation, the Association of Research Libraries<sup>5</sup> recently published a review of current problems in scholarly access as well as suggested remedies entitled *Scholarship, Research Libraries, and Global Publishing*. Building on that report, the Association of Research Libraries and the Association of American Universities<sup>6</sup> inaugurated the Global Resources Program. Its principal goals are to improve access to international research resources and help libraries contain costs through the creation of cooperative structures, the use of new technologies, and the expansion of international document delivery. Funded by a grant from The Andrew W. Mellon Foundation, the Global Resources Program promotes a distributed, interdependent

---

<sup>4</sup>Please see Appendix 9 for more information on double keying and Appendix 6c for a description of SGML. No more than one error in 5,000 characters will be permitted under the subcontract for data entry.

<sup>5</sup>The Association of Research Libraries (ARL) is a not-for-profit membership organization comprising the libraries of 121 North American research institutions and operates as a forum for the exchange of ideas and as an agent for collective action. Membership in ARL is institutional. ARL programs and services promote equitable access to and effective use of recorded knowledge in support of teaching, research, scholarship, and community service.

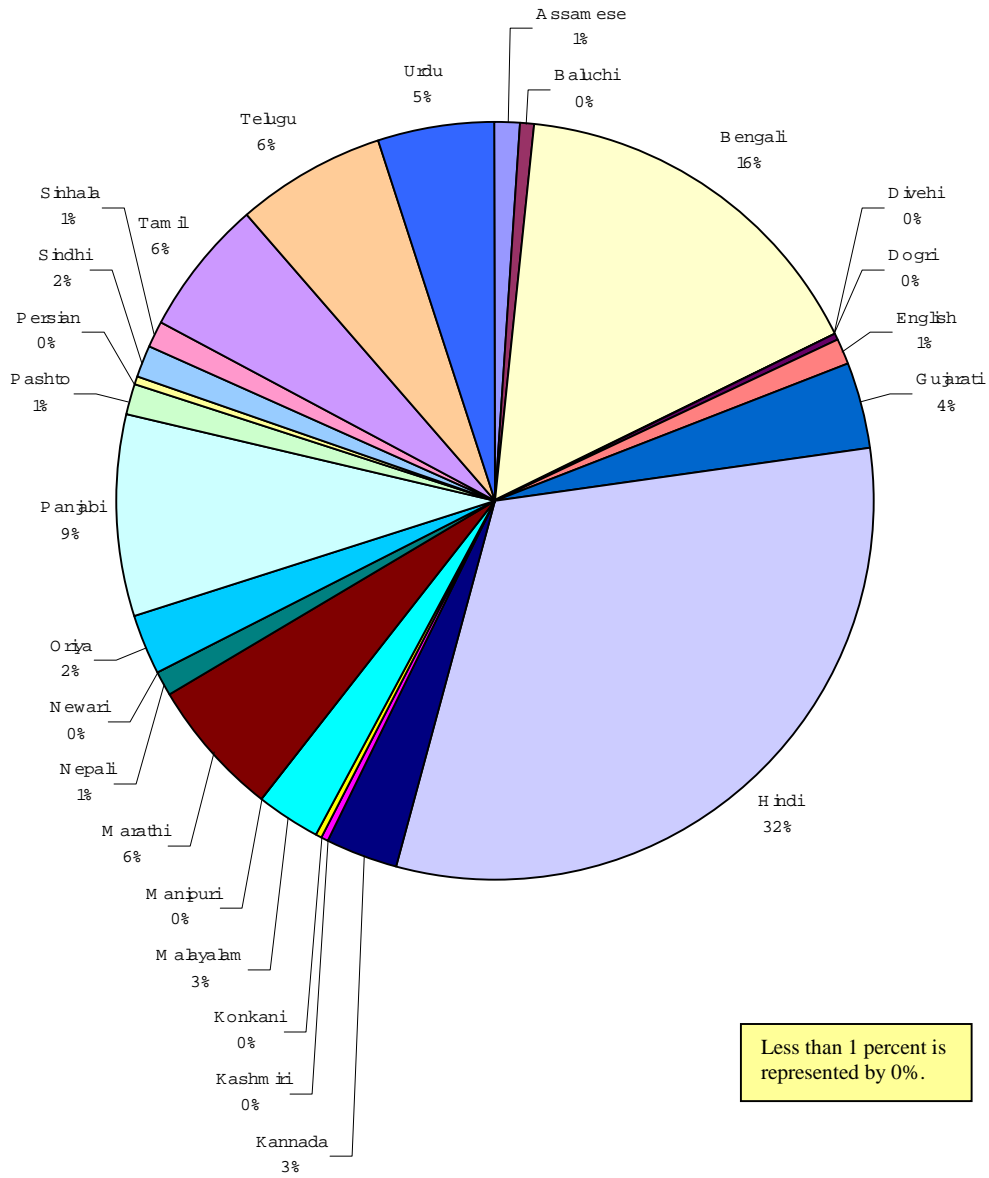
<sup>6</sup>The Association of American Universities (AAU) was founded in 1900 by a group of fourteen universities offering the Ph.D. degree. The AAU currently consists of 60 North American universities. The Association serves its members by assisting them in developing national policy positions on issues that relate to academic research and graduate and professional education. It also provides members with a forum for discussing a broad range of other institutional issues, such as undergraduate education.

approach to collecting scholarly materials from abroad. This approach ensures more focused collection development at individual institutions resulting in access to a broader array of these often difficult-to-acquire resources. The Global Resources Program coordinates its efforts with those of other libraries and organizations that share common interests with the Association of American Universities and the Association of Research Libraries. Currently the Global Resources Program regional projects include: the Cooperative African Newspapers Project; the German Resources Project; the Japan Journal Access Project; the Latin Americanist Research Resources Project; and the Digital South Asia Library. The Digital South Asia Library led by two of the applicants for this proposal, Chicago and Columbia, is the only one funded by the Association of Research Libraries for South Asia. The digital lexicons prepared under the Digital Dictionaries of South Asia project will be integrated with the Digital South Asia Library resources, giving the dictionaries a wider visibility by setting them in a larger context.

The digital dictionaries of South Asia will **serve a large target audience**. The Association for Asian Studies lists more than 760 of its members as having South Asian countries as a major focus of academic interest. Many of these scholars are currently at Title VI South Asia National Resource Centers or have studied at one of the centers. Most of these members have studied at least one South Asian language. In addition, many members of other scholarly associations have a primary interest in South Asia – the American Academy of Religion, the American Political Science Association, the American Anthropological Association, and the American Oriental Society – to name a few. Looking to the larger body of citizens, the 1990 U.S. census shows that slightly more than one million Americans trace their heritage to the cultures of South Asia. Digital Dictionaries of South Asia resources will serve "heritage learners" among this population. Finally, those outside the U.S. will be well served by the dictionaries freely available over the Internet. Many of the printed dictionaries covered by this project are difficult to obtain in South Asia itself. It is important to recall that the South Asian subcontinent encompasses more than 22% of the world's population and that access to the Internet is expanding rapidly in that region.

### Languages of South Asia by Numbers of Speakers

Languages of **Bangladesh**: Bengali, English, Urdu.  
 Languages of **India**: Assamese, Bengali, Dogri, English, Gujarati, Hindi, Kannada, Kashmiri, Konkani, Malayalam, Manipuri, Marathi, Oriya, Panjabi, Tamil, Telugu, Urdu.  
 Languages of the **Maldive Islands**: Divehi, English.  
 Languages of **Nepal**: English, Nepali, Newari.  
 Languages of **Pakistan**: Baluchi, English, Kashmiri, Panjabi, Pashto, Persian, Sindhi, Urdu.  
 Languages of **Sri Lanka**: English, Sinhala, Tamil.



Less than 1 percent is represented by 0%.

|          |         |           |          |         |         |          |       |         |
|----------|---------|-----------|----------|---------|---------|----------|-------|---------|
| Assamese | Baluchi | Bengali   | Divehi   | Dogri   | English | Gujarati | Hindi | Kannada |
| Kashmiri | Konkani | Malayalam | Manipuri | Marathi | Nepali  | Newari   | Oriya | Panjabi |
| Pashto   | Persian | Sindhi    | Sinhala  | Tamil   | Telugu  | Urdu     |       |         |

Examples of use by Digital Dictionaries of South Asia patrons, as identified in Section 2, substantiate the significance of this project. Scholars in virtually all disciplines of the social sciences and humanities will be able to consult this project's Web site for lexical support. One can predict the following scenarios of use:

- Language and literature instructors will use the digital South Asia dictionaries with other electronic texts as an integral part of instruction.<sup>7</sup> Further, the dictionaries will be an asset for those teachers developing Web-based resources for language instruction since they will be available without charge for integration into larger packages of teaching software. The new, high quality pedagogical resources will enable distance language learning by students not able to meet face-to-face with one of the few U.S. instructors for certain least commonly taught languages. These language resources will be valued by "heritage learners," students at the various U.S. Government language schools such as the Defense Languages Institute, as well as others.
- Literary researchers will be able to consult historical examples of word use by prominent litterateurs. The sophisticated search engines deployed under this project will allow readers to extract, for example, all cited uses of verbs by a particular author.
- Linguists will benefit from Web interfaces permitting searches over all the project's South Asian dictionaries or just those from a specific language family, such as the Dravidian family. Further, they will have a solid base of digital data for use in the compilation of new dictionaries.
- Newspaper reporters and foreign business personnel often find Indian uses of English bewildering. The on-line version of *Hobson-Jobson; a glossary of colloquial Anglo-Indian words and phrases, and of kindred terms, etymological, historical, geographical and discursive* will guide them to a more precise interpretation of Indian words and phrases.
- Museum staff will consult the dictionaries for technical meanings of art historical and architectural terms in regional languages.
- School teachers will be able to direct so-called "heritage learners" as well as others interested in South Asian languages to dictionaries created under this project for use in conjunction with other electronic pedagogical resources for language instruction in the less commonly taught languages of South Asia. Further, the National Resource Centers at Chicago, Columbia, and the Triangle South Asia Consortium will offer two summer seminars for high school teachers on use of digital dictionaries created under this project along with other language pedagogy resources.
- Librarians, especially catalogers, are often expected to deal with South Asian texts in languages where they are not fluent. The dictionaries will be readily accessible to assist them and since the Web search engine will permit searching in many more ways than entries in simple head word order, they will be able to find information more easily.

---

<sup>7</sup>For example, Project Madurai at <<http://www.tamil.net/projectmadurai/>> provides many electronic texts in the Tamil language. Further, carefully selected on-line grammars and readers will complement the dictionaries created under this project if our complementary proposal to support their conversion is awarded.

The reach of this project will be further extended through a **matching effort by Columbia University**. Columbia will convert to electronic form a dictionary of the most important classical language, Sanskrit. The Project Director will also seek further funding to support conversion of dictionaries for other classical languages, Pali and Prakrit. Because the vocabulary of many modern Indian languages draws upon the classical languages, these dictionaries will be vital for the larger lexical site on South Asia.

### **Universal Access via the World Wide Web**

Access to the digital South Asia dictionaries will be via a Web site maintained by ARTFL, an acronym for American and French Research on the Treasury of the French Language. ARTFL, based at the University of Chicago, has a reputation among researchers in the humanities for building databases that are easy to use, consistent, and attractive in display and functionality. (Please see the ARTFL Web page for Roget's Thesaurus in Appendix 5.) The project's Web interface will permit searching for dictionary entries or groups of entries based on the following categories, either singly or in combination with one another: head word, meaning, part of speech, and literary citation, to name a few. It will also be possible to search across an entire dictionary database for occurrences of a word or part of a word anywhere in the text.

The project Web site will be a free service on which resources created under this project are accumulated. It will be possible to view documents while connected as well as retrieve documents via conventional file transfer protocols. Programmers will deploy the latest technologies for the project's electronic databases and Web pages. This will include experimentation with voice synthesizers connected to Web browsers, permitting the blind to access the contents of dictionaries on the Internet. This inclusion of sound will likely permit delivery of pronunciation for head words in the dictionaries. We also expect the Web site to be implemented with programming capable of detecting the type of browser used by a remote client and automatically delivering data in a fashion suited to that client.

The University of Chicago will continue to maintain the digital dictionaries of South Asia Web site following the conclusion of this project. As noted elsewhere, the Project Director will seek additional funds to add more digital lexicons to the site.

### **3. Expected Contribution to Mandate under Section 605**

"SEC. 605. RESEARCH; STUDIES; ANNUAL REPORT.

(a) AUTHORIZED ACTIVITIES.—The Secretary may, directly or through grants or contracts, conduct research and studies that contribute to achieving the purposes of this part. Such research and studies may include—

(6) the development and publication of specialized materials for use in foreign language, area studies, and other international fields, or for training foreign language, area, and other international specialists;"

Developing and publishing digital dictionaries of South Asia on the World Wide Web is well within the mandate of Section 605. As demonstrated in the two sections above and elsewhere below, this project will make a material contribution to known needs among scholars from many disciplines, language teachers, students, and a large segment of the general U.S. population. More specifically, the examples of use for these dictionaries provided in Section 2 highlight the major contributions that will be made to language and area study and teaching.

"(b) Annual Report.—The Secretary shall prepare, publish, and announce an annual report listing the books and research materials produced with assistance under this section."

Information provided to the Department of Education during the course of this project will identify the uniform resource location (URL) for dictionaries published on the Internet under this project and the method for ordering copies of the resulting compact disc (CD) containing the dictionary data. This will assist in production of the Department's required annual report.

### **4. Account of Related Materials**

**A Web of On-line Dictionaries** is one of the most comprehensive listings of dictionaries available via the Internet. It is maintained at <<http://www.facstaff.bucknell.edu/rbeard/diction.html>> by Robert Beard of Bucknell University with links to more than 800 electronic dictionaries. Considering the listings at that site for the Indo-Iranian and Dravidian language families,<sup>8</sup> two of the major language families of

---

<sup>8</sup>The URL for Indo-Iranian dictionaries is <<http://www.facstaff.bucknell.edu/rbeard/diction3.html#indic>> and Dravidian is <<http://www.facstaff.bucknell.edu/rbeard/diction2.html#dravidian>>.

South Asia, as noted in Section 1, one finds that all linked dictionaries are either deficient or extremely deficient:

- almost all the dictionaries are rudimentary and many are amateur productions;
- with two or three exceptions, none are based on the best lexical sources;
- few are implemented at an institution with a stable presence on the Web, leading to questions about long-term availability of the resources;
- sophisticated searching is not possible since none are marked with Standard Generalized Markup Language tags; and
- almost none are available for downloading to ones personal computer as a whole file.

In this environment it is clear that the Digital Dictionaries of South Asia project will not duplicate efforts but rather produce electronic lexicons that supersede those now available.

In a **limited number of cases an electronic version will already exist** of a dictionary selected by the project's Advisory Panel. For instance, the Committee on Institutional Cooperation is converting Platts' Hindustani dictionary to a digital file. They are also preparing to mount McGregor's Hindi-English dictionary and two Tamil dictionaries as Web resources. Or as another example, we have recently learned that C.P. Brown's nineteenth-century Telugu-English dictionary was converted to digital form by the University of Hyderabad in India. However, the file does not contain standard markup code nor is it rendered in Unicode. Rather than duplicate efforts or enter the data again, we will seek to collaborate with the Committee on Institutional Cooperation and, in the case of the University of Hyderabad, to augment that data through addition of SGML tags and assist them in mounting the text as a Web resource.

The Project Director and Co-Directors will correspond with the Central Institute of Indian Languages in India, Central Urdu Board in Pakistan, and the Bangla Academy in Bangladesh as this project develops to seek information on other known electronic dictionary projects and to disseminate information about this project in South Asia. When we hear of a project in the subcontinent, we will **seek to collaborate** with them, as described with the University of Hyderabad.

Web based materials can be updated and expanded in a way that materials available on other media cannot. Likewise as compact discs and related format technologies change, items will have to be

"remanufactured" to remain current. That is to say, they are fixed media. But Web-based materials are dynamic and can be updated so that everyone who uses them sees the latest versions.

Finally, the **Digital South Asia Library** is a related resource. This project founded by the University of Chicago and Columbia University is creating new electronic resources for scholarship on the subcontinent. The Digital Dictionaries of South Asia project will fit comfortably within the larger context of the Digital South Asia Library. The Web site featuring periodical index records and full-text of selected titles is located at <<http://www.lib.uchicago.edu/LibInfo/Subjects/SouthAsia/dsal.html>>.

## **5. Likelihood of Achieving Results**

The principal objective of this project is to make available the highest quality electronic dictionaries for South Asian literary languages. Collaboration between Chicago, Columbia, and the Triangle South Asia Consortium will allow a coordinated **division of labor built upon existing institutional strengths** following the methods and procedures described in Section 7 to achieve that objective. Drawing upon a tradition of South Asian scholarship that continues to be supported by significant numbers of faculty and staff as well as considerable material resources, the University of Chicago is already involved in a variety of projects to develop Web-based resources for the study of South Asia. In these projects the expertise of the University of Chicago's American and French Research on the Treasury of the French Language (ARTFL) is a great aid. ARTFL staff have considerable experience in the conversion of texts such as dictionaries into searchable digital files and in use of Unicode for text display over the Internet. Chicago is planning to improve the already extensive Digital South Asia Library begun with Columbia by providing further resources such as the dictionaries described here. Likewise, Columbia University will draw upon its own prestigious tradition of South Asian scholarship and the experience of its faculty and staff with their proven ability to foster institutional cooperation and collaboration to help in the completion of this project. Columbia's SARAI (The World Wide Web Virtual Library for South Asia) has been publicly recognized for its extensive collection of vital South Asian resources available on the Internet. The Triangle South Asia Consortium is building a foundation for South Asian scholarship and has particular experience in the development of technology for language pedagogy. Already a leader in

the televideo delivery of courses it is now ready to build upon that experience by moving forward to delivery over the Internet of course materials for less commonly taught South Asian languages. The dictionaries in this proposal will integrate well with those other language pedagogy materials.

Systems staff of ARTFL will implement the **project's computational and Web delivery** components in the U.S. ARTFL is a University of Chicago program for humanities computing and one of the first large-scale digital libraries developed on the Internet. (Please see Appendix 1b for a detailed statement on the organization and its activities.) ARTFL's involvement will ensure high quality results. Further, it is highly significant that work done under this program will draw upon the body of software previously created by ARTFL thus reducing project expenses.<sup>9</sup> At the same time project funds will support development of new computer algorithms and modes of dissemination over the World Wide Web that will further **augment the software library ARTFL maintains**. The electronic resources created by ARTFL are shared widely throughout academic computing and will therefore serve the needs of a larger audience than this specific project.

The contractor for data entry will likely be the **Input Center**. (Competitive bidding will be the basis for award of contract.) This Chicago firm has an excellent reputation for accuracy of work, timely completion of data entry, and low prices. Both ARTFL and the University of Chicago Library have had successful contracts with the Input Center in the recent past. Conventional double entry techniques are used to ensure very high accuracy, as described in Appendix 9. One specific example is ARTFL's contract with that firm for conversion of Diderot's *Encyclopédie*. Other academic centers have also used

---

<sup>9</sup>The ARTFL software is a search, retrieval, and reporting system for large multimedia databases (texts, images, sound) capable of understanding complex text structures (e.g., SGML, BetaCode) with rich metadata. From its inception ARTFL functions have been and will continue to be designed for scholarly research in databases of literary, religious, philosophical, and historical texts – texts which often serve as the foundation for humanistic research. Important historical encyclopedias and dictionaries have also been ideally suited for digital implementation under the ARTFL system. Many of its most recent projects have involved developing significant new systems that comprise multimedia and hypermedia resources for both teaching and research applications. The ARTFL Project, for example, has digitized Diderot's eighteenth century *Encyclopédie*, a monumental work of 2.5 million words, thousand of plates and legends in a complex system of cross- references. The project represents a full implementation of a hypermedia system supporting full-text retrieval and navigation with hypertextual cross-referencing and full digital imaging support in a single, easy-to-use system. The ARTFL system has and will continue to serve the particular needs of students and scholars in the humanities and a broad range of related fields.

the Input Center. These include the University of California Consortium for Latin Lexicography involving mixed Greek and Roman text and Harvard and Tufts universities' Perseus Project. The Input Center is willing to undertake work in non-roman scripts beyond Greek and have done so successfully in the past. As they state on their Web site at <<http://www.inputcenter.com/>>,

"We are experts in file structure coding. This includes SGML, HTML, Folio, typeset codes, structure codes, image targets and references, and unique linkage codes that can be added to an electronic file to conform to the specific needs of your industry or application. We can also guide you in the construction of custom Document Type Definitions (DTD)."

## 6. Quality of Personnel

The **Project Director** and **Co-Directors** have extensive experience managing complex projects with federal and private funding. This is reflected in the brief curricula vitae included in Appendix 2 for James Nye at the University of Chicago, the Project Director, as well as the two Co-Directors, David Magier at Columbia University and Tony K. Stewart at North Carolina State University, Director of the Triangle South Asia Consortium. Their aggregate experience includes direct engagement with many of the most significant digital library initiatives over the past five years, both at the level of program formulation for national bodies and management of specific projects. They hold positions of leadership on major domestic and international bodies related to area studies scholarship and librarianship. Nye has been at the University of Chicago since 1984, Magier at Columbia University since 1987, and Stewart at North Carolina State University since 1986. All are active in South Asian studies through participation in organizations such as the Association for Asian Studies, the American Oriental Society, the American Institute of Indian Studies, the American Institute of Pakistan Studies, and the American Institute of Bangladesh Studies as well as the library profession through the Committee on South Asian Libraries and Documentation and the American Library Association. Each will commit 10% of his time to the project. There will be no cost to the Federal Government for the time of Nye and Magier whose current work responsibilities include digital library activities. Tony K. Stewart will have a partial course reduction for his extracurricular work organizing and leading the evaluation and planning conferences. The Digital Dictionaries of South Asia project will be well entrusted to their leadership.

A five-member **Advisory Panel** will be constituted to ensure that decisions made during the course of the project are well matched with expectations of users. Professors James Gair at Cornell University, Peter Hook at the University of Michigan, and Colin Masica emeritus at the University of Chicago have agreed to be panelists and have given advice as this proposal was under preparation on the most desirable dictionaries for conversion during the first year. Each is a distinguished South Asia linguist and language teacher. The remaining two panelists will be appointed after the Project Director and Co-Directors consult with the South Asia Council of the Association for Asian Studies.

A part-time **Project Manager** will be appointed by the Project Director in consultation with the Co-Directors. Chief characteristics of the successful candidate will include academic knowledge of South Asian studies, proven skills as an administrator, skills in use of the World Wide Web for dissemination of pedagogical information, and broad exposure to the needs of language learners. The Manager's responsibilities are enumerated in the Plan of Operation in Section 7, below. An average of one-third FTE will be committed over the duration of the project.

**Computer programming** and implementation of Web delivery of the project's resources will be the responsibility of ARTFL. As noted in Section 5 above and in Appendix 1, ARTFL at the University of Chicago is a leader in developing large-scale digital technologies for library initiatives on the World Wide Web. This collaboration extends several years of fruitful engagement between South Asian scholars at Chicago and ARTFL. An average of one-third FTE will be committed over the duration of the project.

**Project assistants** will be recruited from among graduate students in the University of Chicago, Columbia University, and members of the Triangle South Asia Consortium (North Carolina State University, Duke University, the University of North Carolina-Chapel Hill, and North Carolina Central University). Students with advanced training in the regional languages of South Asia will be engaged for work described in Section 7. An average of three-quarter FTE will be committed over the duration of the project.

### **Response to GEPA Section 427**

The statutory standard for equitable access under Section 427 of the Department of Education's General Education Provisions Act is met through free dissemination of the project's resources via the Internet and compact discs as well as through wide publicity of the availability of the project's Web site. As described in greater detail in Appendix 10, the University of Chicago will not discriminate on any basis prohibited by applicable federal, state or local law. All three of the participating universities are equal opportunity employers and provide equitable access to, and participation in, their programs for students, teachers, and others with special needs.

## **7. Plan of Operation**

### **Roles of Project Personnel**

The **Project Director** will have administrative responsibility for the project. Along with his staff at the University of Chicago, he will provide fiscal control, arrange for sub-contracts, and make certain that the program proceeds as described in this proposal. With the **Co-Directors**, he will organize and chair meetings of the project's Advisory Panel, ensure that the quality of data and its presentation on the project's Web site are of high quality, organize two conferences on lexical resources for South Asian studies, and evaluate the project for annual reports and for a concluding project report.

A part-time **Project Manager** will have overall responsibility for the project's smooth operation on a daily basis. This will include work with the Project Director and Co-Directors in coordinating project activities in the U.S., monitoring fiscal records, overseeing subcontracted work with ARTFL and the data entry firm, and supervision of students employed as project assistants.

The **Advisory Panel** will meet annually in conjunction with the South Asia Conference held each October at the University of Wisconsin, Madison. Comments from this body will be central to project management and evaluation of the project both as it progresses and at its conclusion. The Panel will also assist in final selection of the dictionaries for conversion during the project's second and third years.

**ARTFL** will manage the development and implementation of Web resources. Appendix 1 provides a statement of ARTFL's history of success in creation of similar on-line files and maintaining them for use

by scholars and the general public.

Student **project assistants** will assist with resources in the languages and scripts of the region, collaborate with computer programmers from ARTFL, monitor developments related to digital dictionaries on the World Wide Web, connect the project's dictionaries with pedagogical resources for South Asian languages, assist in organizing project conferences, and perform other work as assigned by the Project Manager.

### **Project Design and Plan of Operation**

The plan of operation has several component elements. In chronological order they include:

#### **a) Selection**

The Digital Dictionaries of South Asia project is **designed to select dictionaries most needed** by those identified in Section 1 above. The Advisory Panel (described in Sections 2 and 6) will ensure that the project receives input from a broad range of users and that the dictionaries selected for conversion and dissemination are those most important for the intended audiences. We will solicit input by notices in newsletters to South Asian studies scholars, announcing the digital dictionaries and requesting comments. Further, we will include e-mail response forms on Web pages encouraging readers to recommend other lexicons for consideration under the project. Finally, the evaluation and planning conferences described in Section 9 will be an important means of receiving input.

Appendix 4 includes a **preliminary list of dictionaries** under consideration by the Advisory Panel as this proposal was under preparation. It also includes a table with the languages to be converted each year. Please note that the South Asian languages targeted during the first year are those most frequently taught in this country.

#### **b) Data entry**

We will contract for conversion of the selected dictionaries into digital form. The most widely practiced technique for data entry with high accuracy is **double keying**. This approach involves typists keying in the same text twice. Computer programs determine where the two resulting files are not identical and human editors make corrections as required. Data entry by double keying is described in

Appendix 9. The Input Center has been the successful bidder for several recent contracts with ARTFL and will likely be the contractor for keying the data under this project.

Nine dictionaries – out of a project total of at least thirty-two – will be converted during the first year. This slightly slower pace at the beginning will allow us to test our approach before moving into a higher volume of data entry during the second and third years.

**c) Verification**

It is important to verify that the data delivered by the data entry contractor is as accurate as the contract requires. Specifically, no more than **one error in 5,000 characters** will be permitted. Mr. S. Ramakrishnan, a noted Indian lexicographer, and his staff will undertake data verification for the project in Madras. Using standard statistical sampling techniques, he will determine the error rate in data delivered by data entry contractor. This data verification will take place within two months of the original data entry. Where accuracy falls below the contractually specified levels, the data entry contractor will be instructed to repair the errors.

**d) Dissemination**

The project will be more significant because of the **effective dissemination** of the resulting digital dictionaries. The free project Web site and inexpensive compact discs, as described in Section 11, will allow readers around the world easy access to the on-line lexical reference resources created under this project. This use of the World Wide Web is one of the most effective way of providing equal access to readers. The Co-Directors will notify scholarly organizations, libraries, and the general public of progress during the course of this project by means of news notes and articles in relevant scholarly publications, academic news publications such as the *Chronicle of Higher Education*, postings to the listservers for South Asian scholars, and press releases to national news media. We will send information about this project to all members of the Association for Asian Studies' Committee on South Asian Libraries and Documentation, a group representing all significant North American research libraries with collections on

South Asia.<sup>10</sup> Updates on the project will also be sent on a periodic basis to major foreign libraries.<sup>11</sup>

James Nye, the project's Project Director and also the Director of the University of Chicago South Asia Language and Area Center, will keep other Department of Education National Resource Centers for South Asia informed of progress on this project.

ARTFL implemented computing components of earlier South Asian digital dictionaries at the University of Chicago. Those earlier efforts relied on existing computer programs as well as the extensive knowledge of Mark Olsen, the ARTFL Assistant Director, and his staff of students. Reciprocally, ARTFL staff acquired new expertise in deployment of Unicode characters over the Internet as a result of that collaboration. Those techniques and expertise are now available for this and other future projects.

In our efforts to select and make accessible the best dictionaries, we will **adhere to copyright laws** in the U.S. and in South Asian countries. Fortunately, many of the dictionaries highly valued by scholars were published in the nineteenth century or early in this century. So, they are in the public domain. However, some dictionaries selected by the Advisory Panel are likely to be protected by copyright. The Project Director has experience in negotiating contracts with British and Indian publishers for electronic rights and for permission to convert printed texts to digital format. He will seek the least restrictive terms when negotiating with copyright holders for permission to convert dictionaries under this project. Sometimes publishers respond favorably to an offer that the electronic version of a dictionary will be returned to the publishing house at no cost. On other occasions publishers are impressed with statistical evidence that availability of publications on the Internet improves sales of paper copies rather than

---

<sup>10</sup>Membership includes: Center for Research Libraries, Columbia University, Cornell University, Duke University, Harvard University, Library of Congress, New York Public Library, Syracuse University, and the universities of British Columbia, California at Berkeley, Chicago, Hawaii, Illinois, Iowa, Michigan, Minnesota, Pennsylvania, Texas at Austin, Toronto, Virginia, Washington, and Wisconsin at Madison.

<sup>11</sup>At the minimum this will include: National Library of India, Central Secretariat Library (New Delhi), Nehru Memorial Museum and Library, Indira Gandhi National Centre for the Arts, Shastri Indo-Canadian Institute, Oriental and India Office Collections at the British Library, Bodleian Library, Cambridge University Library, Australian National University, the Südasiens-Institut at Universität Heidelberg, Universität zu Köln, Tokyo University for Foreign Studies, Institute of Developing Economies in Tokyo, and the Library of Congress Field Offices in New Delhi and Islamabad.

diminishing sales. Where discussions with copyright holders are unsuccessful, the project may convert dictionaries that were the second or third choices of the Advisory Panel.

The Digital Dictionaries of South Asia project will organize **summer seminars** with South Asia National Resource Centers outreach staff at the University of Chicago, Columbia University, and the Triangle South Asia Consortium. Two summer seminars for high school teachers will focus on use of digital dictionaries created under this project along with other language pedagogy resources, as noted in Section 2.

The project's design features a **commitment to international standards** as it distributes data. For a more detailed statement of those standards, please see Appendix 6 with extracts from relevant Web sites. Where display of text in non-roman characters is required, this project will use Unicode as the default character set.<sup>12</sup> Further, project staff will report to the Unicode Consortium on how the character sets for South Asian languages could be usefully expanded through the inclusion of conjunct characters for a more felicitous display of texts. All electronic resources delivered over the Internet will contain Dublin Core metadata.<sup>13</sup> The project will use Standard Generalized Markup Language (SGML) as its structured encoding scheme for lexical data.<sup>14</sup>

#### e) **Evaluation**

The last element in the plan of operation is **thorough evaluation of the dictionaries** and the extent to which they meet the needs of readers. Please consult the detailed statement on evaluation contained in Section 9.

---

<sup>12</sup>A basic statement on Unicode taken from <<http://www.unicode.org/>> is included in Appendix 6a. In addition to Unicode, readers will be able to elect delivery in a few other major encoding systems such as Indian Standard Code for Information Interchange (ISCII).

<sup>13</sup>The Dublin Core is a simple set of descriptive elements used to describe electronic resources and aid in their discovery on the World Wide Web. Dublin Core had its origins as a research project at the Online Computer Library Center (OCLC). Over the past four years this initiative has achieved international consensus on the form for encoding metadata about those resources, becoming a de facto standard. A basic statement on Dublin Core taken from <<http://purl.org/DC/>> is included in Appendix 6b.

<sup>14</sup> Standard Generalized Markup Language, commonly referred to by its acronym, SGML, is the international standard for the representation of texts in an electronic form independent of specific systems. A basic statement on SGML is included in Appendix 6c, along with extracts on the Text Encoding Initiative application of SGML taken from <<http://www-tei.uic.edu/orgs/tei/>>.

### Time Schedule

The following table and timeline depict the plan of operation during the three-year project:

---



---

#### SCHEDULE OF PROJECT ACTIVITIES

---

##### Prior to Award

- Preliminary selection of dictionaries for conversion during the first year [Advisory Panel in consultation with Nye, Magier, and Stewart]

##### October 1999 - September 2000

- Select, purchase, and install equipment [Nye, Magier, and Stewart]
- Finalize selection of dictionaries for conversion during first year at the October Advisory Panel meeting [the Advisory Panel in consultation with Nye, Magier, and Stewart]
- Conversion of dictionaries to digital form [data entry contractor]
- Review of converted dictionary data [project staff]
- Finalize selection of dictionaries for conversion during the second and third years at the October Advisory Panel meeting [the Advisory Panel in consultation with Nye, Magier, and Stewart]
- Begin negotiations with publishers for copyright release, if necessary [Nye]
- Publish the project Web pages and converted dictionaries [project staff]
- Publicize the project and resources made available under the project [Nye, Magier, and Stewart]
- Submission of progress project report to Department of Education [Nye, Magier, and Stewart]

##### October 2000 - September 2001

- First evaluation and planning conference of language scholars [Nye, Magier, Stewart, and the Advisory Panel]
- Continue conversion of dictionaries to digital form [data entry contractor]
- Review of converted dictionary data [project staff]
- Publish the converted dictionaries on the Web site and as preliminary CD [project staff]
- Publish CD with dictionaries available to date [project staff]
- Submission of progress project report to Department of Education [Nye, Magier, and Stewart]
- First summer seminar for high school teachers on use of digital dictionaries created under this project along with other language pedagogy resources [National Resource Centers at Chicago, Columbia, and the Triangle South Asia Consortium]

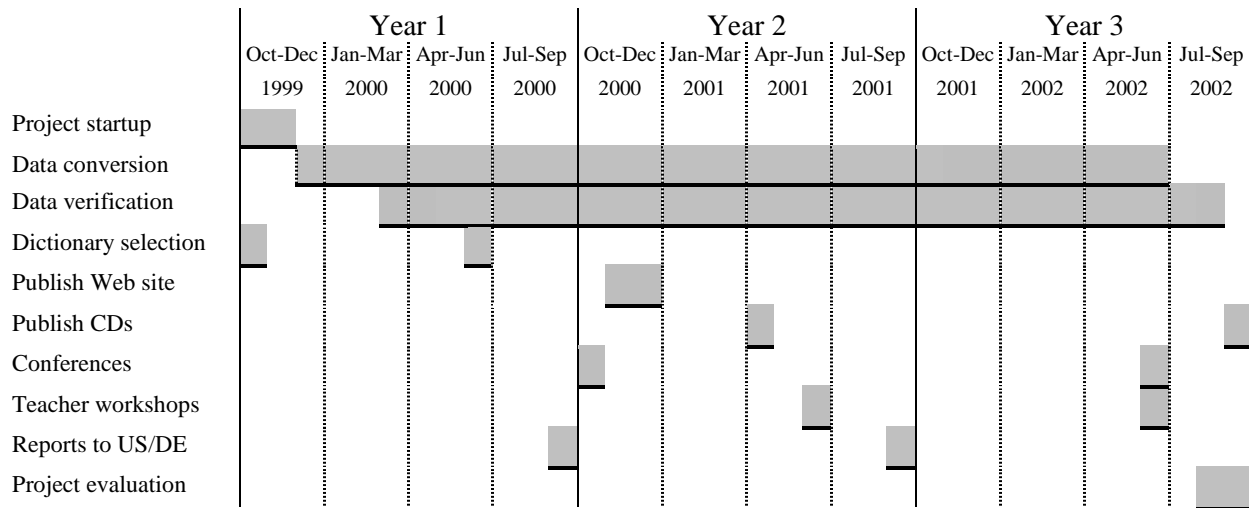
##### October 2001 - September 2002

- Review of converted dictionary data [project staff]
- Publish the converted dictionaries on the Web site [project staff]
- Second evaluation and planning conference of language scholars [Nye, Magier, Stewart, and the Advisory Panel]
- Second summer seminar for high school teachers on use of digital dictionaries created under this project along with other language pedagogy resources [National Resource Centers at Chicago, Columbia, and the Triangle South Asia Consortium]
- Publish complete version of dictionary CD [project staff]
- Formal evaluation of the pilot project [Nye, Magier, Stewart, and the Advisory Panel]
- Submission of final project report to Department of Education [Nye, Magier, and Stewart]

##### Following Conclusion of the Project

- Continuing provision of access to the dictionaries [University of Chicago]

### Project Timeline



### 8. Budget and Cost Effectiveness

This project is designed to be cost-effective and inexpensive relative to its ambitious goals. There are several reasons for expenses being low. First, the project includes more than **50% cost sharing** even though cost sharing is not required under the Title VI International Research and Studies Program. Second, **computer equipment expenses are only three percent** of the total Federal funds requested. Much of the computing equipment to be used is already owned by the three principal universities. Third, a U.S. subcontractor will perform data conversion via the well-attested approach of "off shore" double keying. As noted in Section 7 and Appendix 9, the results are extremely accurate and the prices are low.

The University of Chicago is charging its negotiated rate of indirect costs for this project. However, it is noteworthy that the Dean of the Humanities has agreed to allow 50% of the indirect costs payments to be used for the Project Manager's and student assistant salaries as well as partial payment of data entry expenses.

### 9. Plan of Evaluation

The evaluation of this project will be **systematic and recurring**. Detailed statistics regarding the number of times the Web site is contacted and the nature of that contact will be gathered electronically. In assessing these statistics we will use the widely endorsed "Guidelines for Statistical Measures of Usage

of Web-Based Indexed, Abstracted, and Full Text Resources,"<sup>15</sup> prepared by the International Coalition of Library Consortia, a group currently comprised of over ninety library consortia. Building upon the experience of other projects, these guidelines were developed in order to provide an international standard of preferred practices for the licensing and purchase of electronic information. By using these guidelines we will be able to provide more subtle interpretations of usage statistics than could be achieved by crude counts of "hits" on the project's Web site without violating a user's privacy. The guidelines include queries, menu selections, sessions, items examined and instances of users turned away due to simultaneous user limits. Questions of confidentiality and security are also addressed by the guidelines. In addition to empirical evaluation, comments from patrons will be solicited by means of a response form that will be periodically sent to them electronically. When patrons first register to use the site by providing their e-mail address they will be sent a questionnaire. A further questionnaire will be required on the tenth occasion a reader signs in to use the site. Questionnaires also will be e-mailed to a randomly selected portion of those users who fail to use the site more than once during a year.

The information gathered from statistics on use of the site and patron responses will be considered at **two conferences**. At the first meeting, held in conjunction with the Madison South Asia conference in October 2000 so as to limit expenses, a preliminary evaluation of the project will be produced after a presentation describing the projects goals and demonstrating the resources available at that time. Members of the Advisory Panel will assess the quality of the dictionaries produced; consider comments gathered from recognized scholars as well as other users over the Internet; and examine the costs of preparing and delivering the dictionaries. A summary of this session will be posted on the Digital Dictionaries of South Asia Web site. Potential problems and improvements will be considered at that time and a formal report delivered to the Department of Education. A further report will also be completed at the end of the second year of the project. Subsequently, at a second conference at North Carolina State University in the spring of 2002, a final evaluation of the project will begin and

---

<sup>15</sup> The complete Guidelines, taken from the URL <<http://www.library.yale.edu/consortia>>, are reproduced in Appendix 3 below.

discussions of its implications for the future development of lexical resources and their application to language pedagogy. The Project Director and two Co-Directors will then organize the observations of the Advisory Panel into a final project report for distribution to the Department of Education and members of the Advisory Panel. The final report will also be posted on the project's Web site.

#### **10. Adequacy of Resources**

The adequacy of resources at the University of Chicago, Columbia University, and the Triangle South Asia Consortium are demonstrated in Appendix 1 and the human resources in Section 6 and Appendix 2. In addition to the strengths of the respective South Asian studies programs, it is noteworthy that all three institutions are members of the Digital Library Federation, a program of the Council on Library and Information Resources. The goal of the Digital Library Federation is to establish the conditions necessary for the creation, maintenance, expansion, and preservation of a distributed collection of digital materials accessible to scholars and the wider public.

The three universities participating in this project already have much of the required electronic equipment. The University of Chicago and ARTFL have a robust base of Web equipment already installed. There will be no charge for the use of those facilities. The project will purchase a limited amount of special equipment needed to accomplish the specific objectives of data review and testing under this project. Only three percent of the total Federal funds will be used to purchase computers and printers. Further it is important to note that all three institutions are deeply involved in development of Internet2. The mission of Internet2 as state on its Web site is to "facilitate and coordinate the development, deployment, operation and technology transfer of advanced, network-based applications and network services to further U.S. leadership in research and higher education and accelerate the availability of new services and applications on the Internet."

#### **11. Description of Final Format**

Electronic dictionaries prepared under this project will be **published on the World Wide Web**. Readers will have the option of consulting the lexical databases using a powerful search engine while connected to the site or downloading the dictionary data via standard transfer protocol for use on their

personal computers. Please see Appendix 5 for an example of the format envisioned for the Web interface.

Readers **downloading the data** will have the option of requesting encoding schemes other than Unicode. The Indian Standard Code for Information Interchange (ISCII) and a limited number of other common encoding schemes will be offered.

For those without Internet connections a **compact disc** containing the dictionary data and software for its display, including the facility to display Unicode characters for the various scripts, will be available for distribution at a minimal charge to cover duplication and distribution.

Readers will be encouraged to use the dictionaries in conjunction with Wordbot, a system for simultaneous display of text and a dictionary on a split screen. As they state on their Web page at <<http://www.wordbot.com/wordbot-no.html>>:

Wordbot is an experimental system for automating reference lookup on the web. When you load a page through Wordbot (such as the page you are now viewing), you can click on any word to look up its meaning in a dictionary. Following links in the document works normally, but so that you can look up words even buried in links, Wordbot replaces the original links in a document with a special symbol. Clicking on this symbol will cause the corresponding page to be loaded through Wordbot. Clicking on words in a link, or anywhere else, will cause those words to be looked up.

In order to improve subsequent editions of the dictionaries, the project proposes to take advantage of the potential for interaction provided by the Internet. Comments and suggestions of scholars and users on individual dictionary entries will be gathered for use by future editors in subsequent phases of the project. These comments will be marked with respect to their expertise and experience of the contributor.

## **12. Provisions for Pretesting**

Columbia University and North Carolina State University will pretest the dictionaries in the United States prior to their publication on the project's Web site. The data verification team in India will also pretest the dictionaries under conditions more likely to be encountered by readers in the South Asian subcontinent. Graduate students and others will use the dictionaries through all three means of distribution – via the Web interface, via downloading through file transfer protocol, and on compact disc. They will also use the lexical data in conjunction with Wordbot to test compatibility of the South Asia

dictionaries with that split-screen system mentioned in the previous section. ARTFL programmers informed by the pretesters' critiques will make appropriate revisions to the data and interfaces before the project's products are released to the public.

## **Appendix 1. Institutional Participants**

### **University of Chicago**

The study of South Asia at the University of Chicago has deep historical roots and a strong institutional commitment. Faculty have regularly taught Sanskrit at Chicago since the establishment of the University in 1892. The continuing strength and commitment to South Asian studies at the University of Chicago is reflected by the considerable appropriation of faculty resources, the diversity of languages taught, the established presence of federal Area Studies Center for South Asia, and an intensive level of collection development in the Library.

At the core of the University's strength in South Asian studies is its Department of South Asian Languages where fourteen faculty members dedicate their research and teaching to South Asia. A further twelve faculty members from other departments, including History, Political Science, Anthropology, Art, Linguistics, Behavioral Sciences, and the Divinity School, also focus their research upon South Asia. Together these faculty members comprise the Committee on Southern Asian Studies which provides students with opportunities to pursue graduate degrees with a specialization on South Asia in all of the disciplines described. The consistent excellence of the resources for the study of the subcontinent has been recognized by continuous support for a South Asia National Resource Center from the Department of Education under Title VI and its various predecessors since 1959. This tradition of scholarship is made possible by one of the finest American collections of South Asian material in the United States comprising more than half a million volumes.

The commitment to language instruction is of particular importance for the project outlined here. Along with regular courses in six South Asian languages (Bengali, Hindi, Pali, Sanskrit, Tamil and Urdu), instruction in other area languages is regularly arranged. The expertise of the faculty in language instruction is demonstrated by the fact that many of the most widely used pedagogical resources currently used in the United States were produced by University of Chicago faculty with the support of previous Department of Education grants. These include basic courses as well as advanced readers in Bengali by the late Professor Edward Dimock and similar material for Urdu by Professor C. M. Naim. Both the pedagogical expertise and the linguistic knowledge of these teachers has been and continues to be conveyed to the community of scholars and students at Chicago and will prove an invaluable resource in completing this project.

In addition to the tradition of South Asian scholarship and instruction, the University also has developed a reputation for creative innovation in the use of computing for scholarship. In particular, the ARTFL project has collaborated with scholars in Chicago and elsewhere to provide a number of digital full-text scholarly resources to the academic community. This growing collection includes such databases the University of Toronto's Old English Corpus and Hi-Fido (High Fidelity Digitization On-Line) which dynamically links digital images of sheet music with digital sound recordings of early Chicago jazz. Other projects include collaboration with the Library at the University of Chicago and the Center for Research Libraries in the Electronic Open Stacks project to produce image-based texts as an alternative to microform preservation. In all of these projects ARTFL has developed the expertise, including software, necessary for the kinds of digitization foreseen in this project.

## Appendix 1. Institutional Participants (cont.)

**The ARTFL Project at the University of Chicago**

ARTFL is an acronym for American and French Research on the Treasury of the French Language and is regularly applied to the following three entities: 1) the ARTFL Project which is a cooperative enterprise established in 1981 by the University of Chicago and the French *Centre National de la Recherche Scientifique* (CNRS); 2) the ARTFL database which consists of many thousands of French texts ranging from classic works of literature to non-fiction prose and technical writing originally assembled to prepare a new French historical dictionary; 3) and the ARTFL system developed at the University of Chicago, which is a suite of software for sophisticated indexing, retrieval, and linguistic analysis.

ARTFL as such was among the first large-scale digital libraries developed on the Internet. From its inception in 1981, the ARTFL Project has provided users in North America with exclusively networked-based access to its resources. Well before the appearance of the World Wide Web, ARTFL developed Internet-based client-server access mechanisms (named PhiloLogic). Given ARTFL's experience with Internet-based client-server technology, the Project was able to adapt very quickly to WWW technology as it appeared and began to extend the basic systems to serve the hypertextual and hypermedia capacities afforded by WWW infrastructure. ARTFL provides access to its non-public domain collections through a consortium of 160 subscribing educational institutions in North America. The system handles well over one million queries a month.

The ARTFL Project has a prominent position in the national academic community. Both through its main database and its more recent undertakings such as the development of Diderot and d'Alembert's eighteenth century *Encyclopédie*, ARTFL has become a standard, indispensable reference for scholars and students in a broad range of fields. Results from the ARTFL Project and its technology are featured prominently in a number of important publications, including Keith Baker's *Inventing the French Revolution: Essays on French Political Culture in the Eighteenth Century* (Cambridge, 1990), Liah Greenfeld's *Nationalism : Five Roads to Modernity* (Cambridge, Mass., 1992), Daniel Gordon's *Citizens without Sovereignty: Equality and Sociability in French Thought, 1670-1789* (Princeton, 1994), and Joan DeJean's *Ancients against Moderns: Culture Wars and the Making of a Fin de Siècle* (Chicago, 1997).

The national and international reputation of ARTFL is made manifest by the number and range of its collaborative projects. With the CNRS and scholars at the University of Toronto and Université de Lyon, the ARTFL Project developed electronic versions of the first (1694) and fifth (1798) editions of the *Dictionnaire de l'Académie Française*. The ARTFL system has also been used to treat materials outside of French studies such as the *Opera del vocabolario italiano*, a database of over 1,000 documents in medieval Italian developed in cooperation with the *Consiglio nazionale delle ricerche* in Florence, Italy, the University of Notre Dame, and the University of Reading, England as well as the Swahili-English Dictionary of the Kamusi Project at Yale University and Tamil-English and Hindi-English dictionaries for the CIC at the University of Chicago Library. Also in conjunction with the University of Notre Dame, ARTFL has designed systems to manage large-scale collections of digital images, such as the 12,000 drawings in the Biblioteca Ambrosiana, and interactive video/audio systems for teaching in the *Progetto Italica Corso di Lingua*. Other ARTFL collaborations include work with the Newberry Library, the Center for Research Libraries, the Smart Museum of Art, and professors at George Washington University, Northwestern University, Princeton University, the University of Minnesota, and the University of Ottawa to mount a variety of full-text and multimedia projects.

The address for ARTFL on the World Wide Web is  
<<http://humanities.uchicago.edu/ARTFL/ARTFL.html>>.

### **Columbia University**

Columbia University is recognized nationally as a leader in the development, testing, and implementation of electronic information resources to support teaching and research. The University's strong commitment to promoting innovation in pedagogical applications of technology is seen in a wide range of centrally supported projects, including the Faculty Cluster for Instructional Technologies, the newly created Columbia Center for New Media Teaching and Learning, and the innovative collaborations between the University's Academic Computing, Libraries and the Columbia University Press (which has yielded, for example, the Online Books Project, as well as the popular online full-text resource CIAO - Columbia International Affairs Online). The URL for the Center for New Media is <<http://ccnmtl.columbia.edu/>>. In March of 1999, recognizing the importance and value of these efforts, the University has announced a further strengthening of the human resource infrastructure to support such efforts, by creating a new well-staffed center known as EPIC - Electronic Publishing Initiatives at Columbia. EPIC brings together the best minds and practitioners of electronic publishing from the leader in this area among university press publishers (Columbia University Press), with the researchers, scholars, teachers, students and librarians who represent the wide range of users of such resources, in an effort to centralize the expertise needed to plan and execute effective pedagogically-oriented electronic publications.

Along with its commitment to develop such on-line resources, Columbia has also been very assertive in ensuring broadest possible access to these services. Thus, ColumbiaWeb is available at over 20,000 internet-connected computers in dorm rooms, faculty and department offices, libraries, labs, classrooms, lobbies, student centers, and other locations around campus. Over 40,000 affiliates have individual accounts on the Columbia system, and, together with the wider academic community who access these services over the open World Wide Web, they account for the over 4 million visits per week to ColumbiaWeb.

In concrete terms, this wealth of human and technical infrastructure has made it possible for Columbia Libraries to provide unique and critically needed information services for one of its strongest subject area commitments: South Asian Studies. This is seen, for example, in the creation and development of SARAI (South Asia Resource Access on the Internet), which, because of the quality of its content and its organizational structure, has been designated by the global World Wide Web Consortium (W3C) as the official World Wide Web Virtual Library for South Asia. The URL for SARAI is <<http://www.columbia.edu/cu/libraries/indiv/area/sarai>>. A major digital library Web site, SARAI contains thousands of links to the best of the world's internet resources for South Asian Studies, and is compiled as a collaborative effort with ongoing input from the South Asian Studies community of scholars and librarians. Each link is reviewed and evaluated, annotated, indexed, and classified in a hierarchical structure of linked hypertext sub-collections organized by type, region, country, and subject, following strict bibliographic and selection standards. The author of each resource has been credited. The site also gives direct access to the catalogs of the world's major South Asia library collections and many specialized statistical, bibliographic and full-text resources, including pedagogical resources such as ILM (Inventory of Language Materials), a compendium of South Asian language learning resources compiled and updated continuously online for the last 15 years by Prof. Frances Pritchett (Department of Middle East & Asian Language & Culture, Columbia). SARAI also provides access to the International Directory of South Asia Scholars. This Directory, produced at Columbia and designed to aid scholarly communication and collaboration among South Asianists worldwide, consists of hundreds of full-text entries, written by the scholars themselves, giving contact information and detailed descriptions of their teaching and research projects and interests. The entire directory is keyword searchable for easy access to records by name or by topic or region or language of interest. SARAI is proving to be a very valuable teaching tool and is used in introductory courses on South Asia as well as in pre-dissertation seminars and Internet Tours conducted by the South Asia Librarian. Links to SARAI have been added by the CONSALD librarians at all of the major US Web sites for South Asian Studies programs. Detailed usage statistics from Columbia's Web server, as well as linkages to this site by other libraries, colleges, public schools and organizations, indicate that SARAI is serving a very broad international audience, currently averaging more than 6,000 visits per week. All of the Digital Dictionaries of South Asia proposed in the present project (as well as ancillary materials such as the Digital Library of Lexicography, the Digital Scriptorium of Indic Manuscripts, and the Digital Sanskrit Dictionary produced under the auspices of Columbia's Dharam Hinduja Indic Research Center), will be prominently linked in SARAI for maximum visibility and accessibility.

### **Triangle South Asia Consortium**

The Triangle South Asia Consortium (TSAC) is an educational cooperative of the specialist South Asia faculties of North Carolina State University, Duke University, the University of North Carolina-Chapel Hill, and North Carolina Central University. Formally recognized in 1987, TSAC has grown from a relatively small faculty colloquium into a resource center with a full range of programming and outreach in South Asia. Building on the commitments to area studies on the three major research campuses - the oldest commitments coming from Duke - the Consortium has developed an institutional parity in faculty and programs over the last decade, underscoring its wide administrative support. In recognition of the efforts of faculty at NCCU, a traditionally black institution, the Consortium recently inducted its fourth member. That addition brings to thirty-four the number of South Asia faculty appointments in the Triangle universities. Distributed evenly across the disciplines of the Humanities and Social Sciences, approximately three-quarters of those appointments are senior and tenured.

National recognition of this expertise in South Asia has been demonstrated by the awarding of institutional grants by the U.S. Department of Education, the National Endowment for the Humanities, and the Rockefeller Foundation, and contracts from the Council of American Overseas Research Centers, North Carolina Humanities Council, and various of the community colleges of North Carolina. Commensurate to its overall programming, TSAC universities regularly offer Sanskrit and Persian among the classical languages of South Asia. Instructional programs in two vernaculars - Bengali and Tamil - are under development, while the Hindi and Urdu program has grown to the second largest in the US. The curriculum in Hindi-Urdu languages and literatures shares four teachers through a pioneering and highly successful interactive televideo instruction, developed through a grant from the U.S. Department of Education.

To coordinate its diverse programming, TSAC operates from the NCSU campus, where the secretariat is located. The Executive Director, Afroz Taj, assumes responsibility for cultural programming and public events, including music and dance performance, literary circles, student organizations, and study abroad. The North Carolina Center for South Asia Studies (NCCSAS), which serves as the collaborative research and experimental teaching division of TSAC, is directed by Tony K. Stewart. In this position he also leads development, acts as interlibrary and inter-university liaison, and chairs the Executive Committee. The TSAC Executive Committee, which is composed of three representatives from each of the three original institutions and two from newest member NCCU, is responsible for determining policy and choosing program emphasis, including colloquia, seminars, workshops, research institutes, and team- and televideo-teaching. The location of TSAC's offices in D. H. Hill Library on the NCSU campus signals the strong relationship between the Consortium's programming commitments and the role of the Triangle libraries in collaboration.

Since 1947, TSAC's constituent institutions and the National Humanities Center have been linked through the Triangle Research Library Network (TRLN). Collection holdings now exceed 12 million volumes with only 30% overlap. Duke University anchors the South Asia research collection with more than 200,000 volumes in the vernaculars, but recently acquisitions of South Asia materials in all languages in the TRLN libraries have exceeded \$100,000 per year. Importantly, the institutional commitment to the digital library (NCSU is a member of the Digital Library Federation), coupled with the role of the Triangle's universities in the development of the super high-speed Internet II, places TSAC in a unique position to develop resources for the modern scholar of South Asia.

## Appendix 2. Curricula Vitae for Key Staff

### CURRICULUM VITAE

James H. Nye  
Bibliographer for Southern Asia and Director, South Asia Language and Area Center  
Joseph Regenstein Library  
University of Chicago  
1100 East 57th Street  
Chicago, Illinois 60637 (773) 702-8430

#### Degrees:

- BA Augsburg College, 1968 (philosophy)
- MAR Yale University, 1970 (history of religions)
- MS Drexel University, 1974 (library science)
- [PhD] University of Wisconsin, Madison, comprehensive examinations passed 1982 (South Asian language and literature)

#### Doctoral Dissertation:

Granthamala: A Cultural Study of Indological Publishing in Nineteenth- and Twentieth-Century India  
V. Narayana Rao, Thesis Director

#### Experience:

- 1998-to date, Director, South Asia Language and Area Center, University of Chicago, Chicago, Illinois
- 1984-to date, Bibliographer for Southern Asia, University of Chicago, Chicago, Illinois
- 1974-1984, Librarian II (equivalent to Assistant Professor), Gustavus Adolphus College, St. Peter, Minnesota. Tenured 1979.
- 1978-84, Acquisitions and Reference Librarian
- 1974-78, Reader's Services Librarian
- 1973-to date, Vice President and Editor / Publisher, Musicdata, Inc., Philadelphia, Pennsylvania
- 1971-73, Librarian, Community Legal Services, Philadelphia, Pennsylvania

#### Major Grants, Fellowships, and Honors:

- 1997-99, Andrew W. Mellon Foundation via Association of Research Libraries, grant for "The Digital South Asia Library: A Pilot Project," \$62,700
- 1996-99, Ford Foundation, grant for "Access to Tamil Performance and Folklore Literature," \$250,000
- 1995-97, National Endowment for the Humanities, grant for "Access to Early Twentieth-Century Indian Books: Marathi, Gujarati, Sindhi, Konkani, and English," \$261,835
- 1994-96, U.S. Department of Education, Title II-C, grant for "Tamil Publications for Historical Studies," \$287,690
- 1994-96, National Endowment for the Humanities, grant for "Preservation and Dissemination of Classical and Medieval Tamil Literature," \$198,780
- 1992-94, National Endowment for the Humanities, grant for "Microfilming Nineteenth-Century Hindustani Books," \$140,000
- 1991-94, National Endowment for the Humanities, grant for "Microfilming Nineteenth-Century Hindi Books," \$68,860
- 1991-94, Kern Foundation, grant for "The Adyar Library / University of Chicago Program," \$68,150
- 1989-91, National Endowment for the Humanities, grant for "Preservation of Major Indological Series from the South Asian Subcontinent," \$262,650
- 1989-91, U.S. Department of Education, Title II-C, grant for "Acquisition and Preservation of Indological Series Published in South Asia," \$241,310
- 1985-88, National Endowment for the Humanities, grant for preparation of *South Asian Books in Series: Sanskrit, Pali, and Prakrit*, \$232,000
- 1982-83, Junior Fellowship (dissertation research), American Institute of Indian Studies
- 1981-82, HEA Title VI National Resource Fellowship for Foreign Language Studies, Hindi
- 1974, Beta Phi Mu (national library science honors society)
- 1970 (summer), NDEA Title VI Critical Language Fellowship, Hindi / Urdu

**Selected Publications:**

- A Primer in Oriya Characters* (editor), by Purna Chandra Mishra. New Delhi: Manohar Books, forthcoming 1999.
- South Asian Books in Series: Sanskrit, Prakrit, and Pali* edited by James Nye, associate editor Larry DeVries. Preliminary edition, forthcoming 1999.
- "Recent South Asia Projects at the University of Chicago and the South Asia Microform Project," *South Asia Library Group Newsletter*, no. 40 (January 1993): 7-13.
- "Toward a Sociology of South Asian Book Preservation," in *Planning Modernization and Preservation Programmes for South Asian Libraries*, edited by Kalpana Dasgupta. (National Library conferences, no. 5.) Calcutta: National Library, 1992.
- "Textual Information Retrieval and Analysis for Indology," in *Indological Studies and South Asia Bibliography*, edited by Ashin Das Gupta. (National Library Conferences, no. 3.) Calcutta: National Library, 1988.
- A Primer in Telugu Characters* (editor), by Edward C. Hill. New Delhi: Manohar Books, 1988.
- A Primer in Grantha Characters* (editor and publisher), by K. Venugopalan. 1983.
- Sacred Choral Music in Print and Secular Choral Music in Print* (editor, with T. Nardone and M. Resnick). Philadelphia: Musicdata, 1974.

**Selected Recent Conference Panels, Seminars, and Papers:**

- 1999, "The Unprofitable World of Preservation Micrographics" invited paper, annual meeting of the Association for Asian Studies
- 1997, "Preservation Programs for South Asia" invited paper, annual meeting of the American Library Association
- 1996, "South Asia and the Strategic Plan for Improving Access to Global Information Resources" invited paper for a Presidential session, annual meeting of the Association for Asian Studies
- 1992, Organizer and Chair, panel on "Nineteenth-Century Publishing and the Creation of Indianness" also included my paper on "Sanskrit Series Publishing: A Nineteenth-Century Move into the Hinterlands," 20th Annual Conference on South Asia, University of Wisconsin, Madison.
- 1990, "University of Chicago Program for Preservation of Indological Texts" invited paper, National Library of India, Calcutta, Seminar on Library Modernization and Preservation
- 1989, "Constructing Indianness: Publishing and Culture in South Asia" invited paper, Committee on Southern Asian Studies, University of Chicago
- 1988, "A Machine-Readable Sanskrit Lexicon" invited paper, Sanskrit Database Conference, University of Texas at Austin, Center for Asian Studies
- 1985, "Upapuranas and Mahapuranas: Appendix or Appende?" invited paper, Conference on the Puranas at the University of Wisconsin, Madison, a part of the Festival of India
- 1984, Organizer and Chair, panel on "Computing and Sanskrit Texts" also included my paper on "Computer Photocomposition of Devanagari Texts in the United States and India" at VIth World Sanskrit Conference, Philadelphia

**Memberships and Professional Offices:**

- American Library Association** (Conference panel planning committee, Asia and Africa Section, ACRL, 1989; Area Chairman for South Asia, International Relations Round Table, 1989-1993);
- American Oriental Society** (American Committee on South Asian Manuscripts, 1995-to date);
- American Pakistan Research Organization; Association for Asian Studies** (Committee on South Asian Libraries and Documentation, Executive Committee, 1988-1990; Editor of *South Asia Library Notes and Queries*, 1984-1988); **Microfilming Indian Publications Project**, Steering Committee, 1989-to date; **Nineteenth Century South Asia Short Title Catalog Project**, Steering Committee, 1992-to date; **South Asia Microform Project** (Chair, 1989-1992, 1996-1999; Executive Committee, 1987-1992, 1996-1999; Academic Coordinator for special projects, 1991-to date); **Supervisor for Fulbright Library Interns from India**, six-month internships, 1988, 1989.

**Languages:**

- Natural: Sanskrit, Hindi, Urdu, French, German
- Programming: SNOBOL, PL/1, Revelation database programming

**CURRICULUM VITAE**

David S. Magier  
 South Asia Librarian  
 304 International Affairs  
 Columbia University Libraries  
 New York, NY 10027

212-854-8046  
 212-854-3834 fax  
 magier@columbia.edu

**EDUCATION:**

- Ph.D. (with Distinction) Linguistics, University of California, Berkeley, CA, 1983  
 dissertation title: *Topics in the Grammar of Marwari*  
 M.A. Linguistics, University of California, Berkeley, CA, 1979  
 B.A. (with Distinction) Linguistics, Cornell University, Ithaca, NY, 1977

**PROFESSIONAL EMPLOYMENT:**

- Director of Area Studies*, Columbia University Libraries, NY, 1989 - present. Line responsibility for the 5 departments of the Area Studies Division (staff of 13). Allocating, coordinating and supervising all activities for research-level area studies collections in Judaica and Middle Eastern, Slavic and East European, Latin American and Iberian, South and Southeast Asian, and African studies. Design and delivery of new online information services and networked resources. Serve on LibraryWeb Steering Committee. Liaison with Columbia's extensive international and area studies programs. Serve on Management Council to plan and coordinate the operation of the entire Columbia Libraries system. Work with regional and national library consortia to devise effective library cooperation and coordination programs.
- South and Southeast Asian Studies Librarian*, Columbia University, NY, 1987 - present. Direct responsibility for operation of the South/Southeast Asian Studies Department, and for the assessment, development, maintenance and preservation of the research-level interdisciplinary collections of materials, in all languages and from all regions of the world, on South Asian and Southeast Asian studies, and direct provision of reference, consultation, bibliographic instruction, and electronic information services to these area studies faculty and students. Initiated, designed and maintain at Columbia the officially designated World Wide Web Virtual Library for South Asia (*SARAI: South Asia Resource Access on the Internet*).
- Library Internet Training Consultant*, Metropolitan NY Library Council (METRO), 1994 - present. Design and conduct customized intensive hands-on Internet training workshops to fill the Internet needs and information service development plans of librarians at several hundred public school districts, community college and University libraries and non-profit and government agencies in the New York area. Also provides direct internet training and consulting services for such agencies around the world.
- Adjunct Professor*, Department of Middle East & Asian Languages and Cultures, Columbia University, 1990-1991.
- Assistant Professor*, Department of Linguistics & Lang., Michigan State University, 1986-1987.
- Field Director*, UC Berkeley Urdu Language Program in Pakistan, 1985-1986. Total responsibility for major program of Urdu instruction for American students residing in Lahore, Pakistan.
- Field Director and English Language Consultant*, University Grants Commission, Government of Pakistan, 1984-1985. Developed, implemented and administered English Language Centers and teaching libraries at three universities in Pakistan.
- Lecturer*, Department of Linguistics, UC Berkeley, 1984.

**MAJOR PRESENTATIONS:**

- "The design of web-based pedagogical tools and tutorials," Faculty/Library Symposium, St. John's University, 1999.
- "Collection Development for the LibraryWeb," METRO Symposium on *Building Library Webs: Integrating Online & Web-Based Resources in a Networked Environment*, 1998.
- "The Internet and the Classroom: discovery, evaluation, selection & organization of internet resources in course-based on-line presentations," Faculty Development Day keynote, Mercy College, 1998.
- "Library collections and access: Supporting global expertise," with Deborah Jakubs. HEA-Title VI-Fulbright/Hayes National Policy Conference, 1997. (Published on the web by ARL).
- Annual Seminars on "Research Strategies for Exploiting the Internet," Ford Foundation Workshops on Research & Research Methods in the Developing World, Columbia University, 1996, 1997, and 1998.
- "Digital Libraries: A View from the Third World." 41st Annual International Conference of SALALM (Seminar on Acquisition of Latin American Library Materials), 1996.
- "Electronic Resources and the Future of European Studies," panel on *The Changing Face of European Studies*, WESS (West European Specialists Section), ALA, 1996.
- "Library Collections and the Internet," full-day workshop at Charleston Conference, 1995.

**MAJOR PRESENTATIONS (cont.):**

- "Electronic Resources for the Study of Buddhism," Seminar in Buddhist Studies, 1995.  
 "Cooperative Collection Development: a model for Internet virtual collections," keynote, RLG-METRO Selectors Group (librarians & administrators of Columbia, NYU, NY Public Library, Rutgers), 1995.  
 "Towards a National Agenda for South Asia Teaching and Research Resources," keynote speech, AAS South Asia Council Conference, Madison, 1994.  
 "Developing Virtual Collections: a case study of The South Asia Gopher," Symposium on Electronic Projects for Research & Teaching, 23rd Annual Conf. on South Asia, Madison, 1994.  
 "The process and value of library networking and coordination," Annual Meeting of Library Directors of DELNET (New Delhi consortium of academic and research libraries), 1994.  
 "Cooperative collection development in the context of competition," Conference on the Future of South Asian Studies, Madison, 1993.  
 "Library Issues: Budgets and Coordination of Acquisitions," a report to the Working Group on Library Issues (David Magier, Chair), US Dept. of Education, Meeting of Title VI Area Studies National Resource Center Directors, Washington, D.C. April, 1993.  
 "Electronic Resources for Humanities Research," lecture/workshop, JvNCNet Internet Resources and Applications Symposium at Princeton University, 1993.

**NATIONAL COMMITTEES, PROJECTS, MEMBERSHIPS:**

- Chair, SAC-East (East Coast S. Asia Library Consortium: Columbia, NY Public Library, U. Pennsylvania, Cornell, Harvard, Syracuse, U. North Carolina, U. Virginia, Library of Congress)  
 Chair, Asian Studies Online Library (an international collaborative project under AAS)  
 Chair (former), Committee on South Asian Libraries and Documentation (CONSALD)  
 Chair (former), South Asian Microfilm Project (SAMP)  
 South East Asian Microform Project (SEAM)  
 Committee on Research Materials on South East Asia (CORMOSEA)  
 Chair, CORMOSEA Task Force on Southeast Asian Network Resources  
 Committee on Publications, Board of Directors, Association for Asian Studies  
 Bibliography of Asian Studies Advisory Committee  
 Executive Committee, Rajasthan Studies Group  
 Association for Asian Studies  
 Board of Trustees, American Institute of Pakistan Studies (AIPS)  
 Committee on Publications, American Institute of Indian Studies (AIIS)

**FELLOWSHIPS AND AWARDS:**

- American Institute of Pakistan Studies grant for field research in Lahore, 1985-1986  
 Fulbright-Hays Doctoral Dissertation Research Abroad grant for field research in India, 1981-1982  
 Berkeley Urdu Language Program in Pakistan fellowship, Lahore, Pakistan, 1979-1980  
 Foreign Language Area Studies Fellowships (Hindi and Urdu) 1978-1983

**SELECTED PUBLICATIONS:**

- "Reflections of an early explorer," in Patricia Libutti (ed.) [1999] Librarians as Learners, Librarians as Teachers: Diffusion of Internet Expertise in the Academic Library. Chicago: ARCL. January 1999.  
 "The language of the bard," in Ann G. Gold, [1992] A carnival of parting : the tales of King Bharthari and King Gopi Chand as sung and told by Madhu Natisar Nath of Ghatiyali, Rajasthan. Berkeley: University of California Press, 1992. Pp. 335-350.  
 "Dative/accusative subjects in Marwari," in Manindra Verma, K. P. Mohanan (eds.) [1990] Experiencer Subjects in South Asian Languages. Stanford: Center for the Study of Language & Information.  
 "The Transitivity Prototype: Evidence from Hindi," in Word 38.3:187-199 (1987).  
The Semantics of Participant Roles: South Asia and Adjacent Areas. Edited by Arlene Zide, David Magier and Eric Schiller. Bloomington: Indiana University Linguistics Club. 1985.  
 "Components of Ergativity in Marwari," in Papers from the Nineteenth Regional Meeting of the Chicago Linguistic Society, April 1983.  
 "Marwari Honorifics," in P.J. Mistry (ed.) Studies in South Asian Languages and Linguistics, South Asian Review 6.3:160-173 (July 1982). South Asian Literary Assoc. (MLA).  
Hindi Films for Instructional Purposes [3-volume textbook and materials for teaching Hindi], with Bruce Pray, Satti Khanna. UC Berkeley: Center for South & Southeast Asia Studies, 1980.

**LANGUAGES:**

Hindi, Urdu, Rajasthani, Spanish

## Appendix 2. Curricula Vitae for Key Staff (cont.)

**CURRICULUM VITAE****Tony K. Stewart**

Department of Philosophy and Religion  
010 Winston Hall, Box 8103  
North Carolina State University  
Raleigh, NC 27695-8103

Triangle South Asia Consortium  
3220 D. H. Hill Library, Box 7111  
North Carolina State University  
Raleigh, NC 27695-7111

***University Education***

Ph.D. with Distinction

Department of South Asian Languages and Civilizations, The University of Chicago (1981-1985)

Dissertation Title: "The Biographical Images of Krsna-Caitanya: A Study in the Perception of Divinity"

A.M.

Department of South Asian Languages and Civilizations, The University of Chicago (1976-81)

B.A. Summa Cum Laude (Honors)

Major: Religious Studies, Department of Philosophy and Religion

Minors: German, Asian Studies

Western Kentucky University (1972-76)

***Professional Experience***

Director

North Carolina Center for South Asia Studies, Triangle South Asia Consortium

1998-present

Adjunct Associate Professor of South Asian Religions

Department of Religious Studies, The University of North Carolina at Chapel Hill

1993-present

Associate Professor of History of Religions

Department of Philosophy and Religion, North Carolina State University

1992-present

Director

Triangle South Asia Consortium (an educational cooperative of North Carolina State University, Duke University, and the University of North Carolina-Chapel Hill)

1987-1996

Assistant Professor of History of Religions

Department of Philosophy and Religion, North Carolina State University

1986-92

***Languages:*** Bengali/Bangla, Sanskrit, French, German

***Select Publications since 1991***

"Surprising Bedfellows: Vaisnava and Shi'i Alliance in Kavi Aripa's 'Tale of Lalamon'." *International Journal of Hindu Studies*. Special issue on Hinduism and Islam, forthcoming.

and Robin Rinehart, "The Anonymous *Agama Prakasa*: Preface to a 19th c. Gujarati Polemic" In *Tantra in Practice*, edited by David Gordon White. Princeton: Princeton University Press, forthcoming.

"The Tale of the Faithful Wife Rambhavati" and "Cosmogony of Ali Roja's *Agama*." In *South Asian Islam in Practice*, edited by Carl W. Ernst. Princeton: Princeton University Press, forthcoming.

and Edward C. Dimock, Jr., "Krttibasa's Apophatic Critique of Rama's Kingsh" in *Questioning Ramayanas*, edited by Paula Richman (Berkeley: University of California Press, in press).

"Alternate Structures of Authority: Satya Pir on the Frontiers of Bengal." In *Beyond Turk and Hindu: Shaping Islamicate Identity in South Asia 1400-1800*. Edited by David Gilmartin and Bruce B. Lawrence. Gainesville: University of Florida, forthcoming.

**Select Publications (continued)**

- "When Rahu Devours the Moon: The Myth of the Birth of Krsna Caitanya." *International Journal of Hindu Studies* 1, no. 2 (August 1997): 1-44.
- "Syncretism" (with Carl W. Ernst) and "Satya Pir." *South Asian Folklore: An Encyclopedia*. Edited by Margaret A. Mills and Peter J. Claus. Garland Publishing Co., forthcoming.
- and Rebecca Manring, "In the Name of Devotion: Acyutarana Caudhuri and the Hagiographies of Advaitacarya," *Journal of Vaisnava Studies*, [special issue on Gaudiya Vaisnavism edited by Joseph T. O'Connell], 5, no. 1 (Winter 1996-97): 103-26.
- "Satya pīrer panca punthi: bikalpita kahinira aitiḥya" (Five Texts of Satya Pir: Alternative Narrative Traditions). Introduction by Tony K. Stewart, edited with notes by Md. Shah Jahan Miya and Tony K. Stewart. *Sahitya Patrika* (forthcoming). Includes: *Tinlaksyapīrer pancali* by Anonymous [Harinarayana], *Trilakhya pīrera pancali* by Harinarayana, *Galakataphasyara pala* by Rasamaya, *Satya pīrer pancali (Vallabha pala)* by Krsnasankara, and *Satyanarayana katha (Aksoti pala)* by Kavi Kanva. In Bengali.
- "Encountering the Smallpox Goddess: The Auspicious Song of Sitala" (389-98), "The Exemplary Devotion of the 'Servant of Hari'" (564-77), "The Goddess Sasthi Protects Children" (352-66), "The Rescue of Two Drunkards" (375-88), "Satya Pir: Muslim Holy Man and Hindu God" (578-97). In *The Religions of South Asia in Practice*, edited by Donald S. Lopez, Jr. Princeton: Princeton University Press, 1995.
- The 'Caitanya Caritamṛta' of Kṛsnadasa Kavirāja*. Translated with commentary by Edward C. Dimock, Jr. Edited by Tony K. Stewart. With an introduction by the translator and the editor. Harvard Oriental Series no. 52. Cambridge, MA: Harvard University Press, in press.
- "One Text from Many: The *Caitanya caritamṛta* as 'Classic' and 'Commentary'." In *According to Tradition: Hagiographical Writing in India*. Edited by Winand M. Callewaert and Rupert Snell. Wiesbaden: Otto Harrosowitz, 1994, 317-56.
- "When Biographical Narratives Disagree: The Death of Krsna Caitanya." *Numen* 38, no. 2 (1991): 231-60.
- "The Biographies of Sri Caitanya and the Literature of the Gaudiya Vaisnavas." In *Vaisnavism: Contemporary Scholars Discuss the Gaudiya Tradition*, edited by Steven J. Rosen, forward by Edward C. Dimock, Jr. New York: Folk Books, 1992, 101-125.
- With Hena Basu. "*Baḥḥaya prakasita vaisnava samayika patrika: ekti tathya nirdesika suci*" (an annotated reference guide to Vaisnava journals in Bengali). *Bangiya Sahitya Parisat Patrika* 90, no. 2 (Vaisakha 1390 BS): 1-8.

**Select Honors and Awards (excluding NCSU grants) since 1990**

- Jubilee Fellow, South Asia Regional Studies, University of Pennsylvania (1997)
- National Endowment for the Humanities, Senior Research Fellowship for University Teachers (1996-97)
- Fellow, Carolina Seminar for Comparative Islamic Societies, University of North Carolina-Chapel Hill (1994-95; 1995-96; 1996-97)
- William J. Fulbright Senior Research Fellowship, CIES (Dhaka, Bangladesh, 1991-92)
- Senior Research Fellow, American Institute of Bangladesh Studies (Dhaka, Bangladesh, 1991-92)
- National Endowment for the Humanities Travel to Collections Grant (London, UK, 1991-92)
- University Outstanding Teacher and inductee into the Academy of Outstanding Teachers, NCSU (1989-90)

**Institutional Grants, National Offices**

- Director, "Integrating South Asia Curricula through Core Course and Shared Language Instruction in the Research Triangle Universities of North Carolina," U.S. Department of Education, Center for International Education, Undergraduate International Studies and Foreign Language Program (1995-98).
- Collaborative Fellow, "South Asian Islam and the Greater Muslim World," Rockefeller Foundation Residency Program, Humanities Fellowships, Arts and Humanities Division, The Rockefeller Foundation; with the Triangle South Asia Consortium (1994-97)
- Co-Director (with Carl W. Ernst, University of North Carolina-Chapel Hill), "Hindu and Muslim: Rethinking Religious Boundaries in South Asia," National Endowment for the Humanities (NEH) Summer Seminar for College Teachers (Summer 1995)
- American Institute of Bangladesh Studies (Treasurer 1992-93, Trustee 1990-present)
- American Institute of Indian Studies (Secretary 1991-95, Trustee 1989-1996)
- South Asia Microform Project (Executive Committee 1995-1997, Trustee 1996-present)

Appendix 3. Letters of Commitment

Columbia University in the City of New York | New York, N.Y. 10027

THE LIBRARIES

535 West 114th Street

Area Studies Library Services  
304 International Affairs  
420 West 118th Street  
New York, NY 10027-7296

212-854-8046  
FAX: 212-854-3834

3/31/99

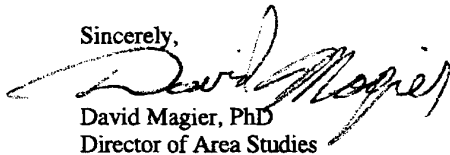
James H. Nye  
(Project Director, Digital Dictionaries of South Asia)  
Director, South Asia Language & Area Center  
5848 University Avenue, Kelly 313  
University of Chicago  
Chicago, IL 60637

BY FAX: 773/753-0569

Dear Jim,

This memo certifies the commitment of Columbia University Libraries to the project described in our joint application to the US Department of Education's International Research and Studies Program, under the title of "Digital Dictionaries of South Asia". As described in that proposal, we commit 10% of my time to the project as Co-Director, for the duration of the project. We commit to the production of a digital Sanskrit Dictionary and a Digital Library of Lexicography in support of the goals of this project. Columbia Libraries also commits to full participation in all the phases and aspects of the project, including allocation of staff, workspace, equipment, data and telecommunications facilities, supervision, maintenance and dissemination of digital material on its webservers, and all other activities as described in the proposal or that may arise in the course of carrying out the project. We make these commitments in recognition of the value of the proposed project both to our own South Asia students and faculty and to the field of South Asian Studies and international education in general.

Sincerely,



David Magier, PhD  
Director of Area Studies  
and  
South Asia Librarian  
Columbia University Libraries

North Carolina State University is a land-grant university and a constituent institution of The University of North Carolina

**Office of Research**

**NC STATE UNIVERSITY**

College of Humanities and  
Social Sciences  
Campus Box 8101  
Raleigh, NC 27695-8101

919.513.1837  
919.515.7856 (fax)

March 30, 1999

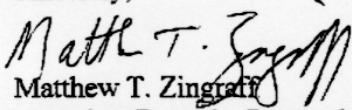
Mr. James Nye  
Regenstein Library  
1100 E.57<sup>th</sup> Street  
University of Chicago  
Chicago, IL 60637

Dear Mr. Nye,

North Carolina State University's College of Humanities and Social Sciences is pleased to participate with the University of Chicago and Columbia University on the three-year "Digital Dictionaries of South Asia" project.

We are committed to this project because of its importance to developing the larger digital infrastructure that will play an increasingly significant role in the academy, government, and industry. To do this, we are prepared on behalf of the Triangle South Asia Consortium to support the project by providing extensive field testing of the dictionaries, and to organize the workshops that will determine the shape of the product and its dissemination.

Sincerely,

  
Matthew T. Zingraff  
Associate Dean for Research

From

S Ramakrishnan,  
Cre-A, 12, 4<sup>th</sup> Cross Street, Karpagam Gardens, Adyar, Chennai 600 020  
Phone : 490 36 67 E-mail: [crea@md3.vsnl.net.in](mailto:crea@md3.vsnl.net.in) Fax: 044- 8284457

To

Mr James Nye  
South Asian Bibliographer, University of Chicago Library, 1100 East 57<sup>th</sup>  
Street, Chicago 60637 IL., U S A Fax: 773-753-0569

Dear Jim,

Please take this letter as a statement of my intention to undertake verification of lexical data for your project titled Digital Dictionaries of South Asia. The work of the project over three years will make an important contribution to scholarship on South Asia. I wish you the best with your application to the U S Department of Education.

Sincerely,



S Ramakrishnan

29.3.99

## Appendix 4. Dictionaries to be Converted

## Preliminary List of Dictionaries

| First Year  | Second Year  | Third Year  |
|---|--|---|
| Comparative<br>Bengali<br>English<br>Hindi<br>Sanskrit (with matching funds)<br>Tamil<br>Urdu | Gujarati<br>Kannada<br>Malayalam<br>Marathi<br>Nepali<br>Oriya<br>Panjabi<br>Sinhala<br>Telugu | Assamese<br>Baluchi<br>Divehi<br>Dogri<br>Kashmiri<br>Konkani<br>Maithili (Hindi dialect)<br>Manipuri<br>Newari<br>Pali (matching funds sought)<br>Pashto<br>Persian<br>Prakrit (matching funds sought)<br>Rajasthani (Hindi dialect)<br>Sindhi |

Languages to be covered, in alphabetical order, are: Assamese, Baluchi, Bengali, Divehi, Dogri, English, Gujarati, Hindi, Kannada, Kashmiri, Konkani, Maithili, Malayalam, Manipuri, Marathi, Nepali, Newari, Oriya, Pali, Panjabi, Pashto, Persian, Prakrit, Rajasthani, Sanskrit, Sindhi, Sinhala, Tamil, Telugu, Urdu.

The Advisory Panel will make final decisions on which dictionaries listed below will be converted.

Dictionaries preliminarily selected for conversion during the first year are marked with a **double vertical bar to the left**.

**ASSAMESE**

Barua, Hemchandra, d.1896.

*The Assamese-English dictionary Hemakosha : etymological Assamese words and idiomatical phrases done into English.* 2d. ed. enlarged. Jorhat, Assam : Barakataki company, 1941. 1030 (i.e. 1038) p. : ill. ; 25 cm.

**BENGALI**

Bandyopadhyaya, Haricharan.

*Bangiya sabdakosha.* [2nd ed.] Niu Dilli : Sahita Akademi, 1966 1967. 2 v. ; 25 cm.

Biswas, Sailendra, 1918-

*Samsad Bengali-English dictionary.* Calcutta, Sahitya Samsad; [distributors: Indian Book Distributing Co.], 1968. 1278 p. ; 19 cm.

Dabbs, Jack Autrey.

*A short Bengali-English, English-Bengali dictionary.* [College Station] Dept. of Modern Languages, A. & M. College of Texas, 1962. xii, 173 p. ; 22 cm.

Dasa, Jnanendramohana, 1871-1939.

*Bangala bhashara abhidhana.* 1917. 2 v. ; 27 cm.

Dev, Ashu Tosh.

*Dev's concise dictionary (English to Bengali & English).* Rev. ed. [enl. with appendices.] Calcutta, S. C. Mazumder, 1958. 784 p. ; 20 cm.

Sen, Sukumar.

*An etymological dictionary of Bengali, c. 1000-1800 A.D.* Calcutta, Eastern Publishers 1971. 2 v. (xv, 968 p.) ; 24 cm.

Shahidullah, Muhammad, ed.

*Purba Pakistani ancilika bhashara abhidhana.* Dhaka, Banla Ekademi, 1965. 3 v. 28 cm. (Purba Pakistani Bamlara adarsa abhidhana, khanda 1-3) Bengali dialects.

**ENGLISH**

Wilson, H. H. (Horace Hayman), 1786-1860.

*A glossary of judicial and revenue terms, and of useful words occurring in official documents relating to the administration of the government of British India, from the Arabic, Persian, Hindustani, Sanskrit, Hindi, Bengali, Uriya, Marathi, Guzarathi, Telugu, Karnata, Tamil, Malayalam, and other languages, compiled and published under the authority of the Honorable the Court of Directors of the East-India Company.* [2d ed.] Delhi, Munshiram Manoharlal [1968] xxiv, 727 p. 29 cm. (First ed. 1855.)

## Appendix 4. Dictionaries to be Converted (cont.)

Yule, Henry, Sir, 1820-1889

*Hobson-Jobson; a glossary of colloquial Anglo-Indian words and phrases, and of kindred terms, etymological, historical, geographical and discursive.* New ed. edited by William Crooke, B.A. London, J. Murray, 1903. xlviii, 1021 p. 23 cm.

**GUJARATI**

Belsare, Malhar Bhikaji.

*An etymological Gujarati-English dictionary.* 2nd ed. rev. & enl. Ahmedabad : H.K. Pathak, 1904. xii, 1207 p. ; 25 cm.

**HINDI AND URDU**

*Bhargava's standard illustrated dictionary of the Hindi language (Hindi-English edition).* Rev. and enl. Varanasi, N. Bhargava 1946. 1280 p. : ill.

Caturvedi, Mahendra.

*A practical Hindi-English dictionary.* Delhi, National Pub. House, 1970. 16, 738 p. ; 23 cm.

Fallon, S. W., 1817-1880.

*A new Hindustani-English dictionary, with illustrations from Hindustani literature and folk-lore.* Banaras, London, Printed at the Medical hall press; Trübner and co., 1879. 3 p. l., xxiv p., 2 l., 1216, ix p. ; 26 cm.

*Ferozsons Urdu-English dictionary : Urdu words, phrases and idioms with English meanings and synonyms.* Lahore : Ferozsons, 1960. 2, 831 p. ; 26 cm.

Hoernle, A. F. Rudolf (August Friedrich Rudolf), 1841-1918.

*A Comparative dictionary of the Bihari language.* Calcutta, Bengal Secretariat Press, 1885- v. : col. maps. ; 32 cm.

Lalasa, Sitaramma, ed.

*Rajasthanni sabada kosa; Rajasthanni Hindi brhat kosa.* [1. samskarana] Jodhapura : Rajasthanni Sodha Samsthana, 1962- <v. 1; v. 2, pts. 1-2; v. 3, pts. 1-3; v. 4, pts. 1-3 in 9 > ; 28 cm.

Mallik, Bhaktiprasad.

*Dictionary of the underworld argot : West Bengal and the Bhojpuri and Magahi areas of Bihar.* Calcutta : Praci Bhasha Vijnan, Indian Journal of Linguistics, 1976- v. ; 22 cm.

*Manaka Hindi Kosa; Hindi bhasa ka adyatana, artha-pradhana aura sarvangapurna sabdakosa.* [1. samskarana] Prayaga, Hindi Sahitya Sammelana, 1962 1966. 5 v. ; 28 cm.

Platts, John Thompson, 1830-1904.

*A dictionary of Urdu, classical Hindi, and English.* London, W.H. Allen & Co., 1884. viii, 1259, [1] p. ; 27 cm.

**KANNADA**

Kittel, F. (Ferdinand)

*Kittel's Kannada-English dictionary.* Rev. and enl. by M. Mariappa Bhat. [Madras] University of Madras, 1968-1971. 4 v. port. 29 cm.

"First edition published by Basel Mission Book and Tract Depository, Mangalore, 1894."

**KASHMIRI**

Grierson, George Abraham, Sir, 1851-1941.

*A dictionary of the Kashmiri language.* Calcutta, Asiatic society of Bengal, 1932. xxiii, [1], 1252 p. (Bibliotheca Indica, 229)

**MALAYALAM**

Bailey, Benjamin, 1791-1871.

Bancamin Beyliyute Malayalam-Inglish nighantu. [Kottayam, Gurunadhan Printers & Publications, 1970. 21, 930 p. : illus. ; 26 cm.

Gurunadhan publication no. 12. Half title: Malayalam-Inglish nighantu. Original t. p. reads: A dictionary of high and colloquial Malayalam & English ... by B. Bailey. Cottayam, Printed at the Church Mission Press, 1846.

Gundert, Hermann, 1814-1893.

*A Malayalam and English dictionary.* Mangalore, London, C. Stolz; Trübner & co. [etc., etc.] 1872. xviii, 1116 p. ; 24 cm.

Kunjan Pillai, Suranad.

*Malayalam lexicon : a comprehensive Malayalam--Malayalam-English dictionary on historical and philological principles.* Trivandrum : London : University of Kerala ; Distributed by Books from India, 1965- v. ; 29 cm.

Vol. 1 has added title page in Malayalam: Malayala mahanighantu. Vol. 3 edited by K. V. Nambudiripad. English and Malayalam.

**MARATHI**

Deshpande, M. K. (Madhav Kashinath), 1910- comp.

*Marathi-English dictionary.* Poona : Suvichar Prakashan Mandal, 1967. 604 p. ; 20 cm.

Molesworth, J. T. (James Thomas), 1795-1872.

*A dictionary, Marḥṭhi and English.* 2d ed., rev. and enl. by J. T. Molesworth. Bombay, Printed for government at the Bombay Education Society's press, 1857. xxx, 920 p., 1 l. ; 33 x 25 cm.

## Appendix 4. Dictionaries to be Converted (cont.)

Vaze, Shridhar Ganesh.

*The Aryabhushan school dictionary, Marathi-English.*  
Poona, Arya-Bhushan press, 1911. [8], 576 p.

---

**NEPALI**

Rana, Gabriel.

*Nepali-English dictionary.* Darjeeling, Shyam Bros.  
1968. 302 p. ; 19 cm.

---

Turner, R. L. (Ralph Lilley), 1888-

*A comparative and etymological dictionary of the Nepali language.* London, K. Paul, Trench, Trubner, 1931.  
xxiii, 935 p. ; 32 x 26 cm.

---

**NEWARI**

Hale, Margrit, 1934-

*A vocabulary of the Newari language.* Kirtipur,  
Summer Institute of Linguistics [and] Institute of Nepal  
Studies, Tribhuvan University, 1971. 40, 40 l. ; 30 cm.  
(Comparative vocabularies of languages of Nepal)

---

Manandhar, Thakur Lal.

*Newari-English dictionary : modern language of Kathmandu Valley.* Delhi : Agam Kala Prakashan, 1986.  
xlix, 284 p. ; 29 cm.  
(École française d'Extreme-Orient produced the dictionary.)

---

**ORIYA**

Brooks, William.

*An Oriya and English dictionary, designed for the use of European and native students and schools.* Cuttack,  
Orissa Mission press, 1875. [2], 259 p.

---

**PALI**

Pali Text Society, London.

*The Pali Text Society's Pali-English dictionary.*  
Chipstead, 1921-1925. 8 v. in 1.

---

**PANJABI**

Newton, J.

*A dictionary of the Panjāb language.* Lodiana,  
Printed at the mission press, 1854. vi, 438 p.

---

Singh, Maya.

*The Panjāb dictionary.* Lahore, Munshi Gulab Singh  
& sons, 1895. 2 p. l., 1221 p. ; 28 cm.

---

**PASHTO**

Bellew, H. W. (Henry Walter), 1834-1892.

*A dictionary of the Pukhto or Puksho language, in which the words are traced to their sources in the Indian and Persian languages.* London : Thacker & Co., 1867. xi, [1], 355 p. ; 26 x 20 cm.

Puksho and English.--English and Puksho.

---

**PERSIAN**

Hayyim, Sulayman.

*New Persian-English dictionary : complete and modern, designed to give the English meanings of over 50,000 words, terms, idioms, and proverbs in the Persian language, as well as the transliteration of the words in English characters.* Teheran : Beroukhim, 1929-1931, 2 v. ; 22 cm.

---

Steingass, Francis Joseph, 1825-1903.

*A comprehensive Persian-English dictionary, including the Arabic words and phrases to be met with in Persian literature.* London, Routledge & K. Paul, 1892.  
viii, 1539 p. ; 28 cm.

---

**PRAKRIT**

Ippagumta, S.

*Camdu kosam : Prakrit:English dictionary.* Delhi,  
India : Parimal Publications, 1992. xii, 217 p. ; 23 cm.

---

**SANSKRIT**

Apte, Vaman Shivaram, 1858-1892.

*Revised and enlarged edition of Prin. V. S. Apte's The practical Sanskrit-English dictionary.* Poona : Prasad Prakashan, 1957- v. <1-2 > ; 28 cm.

---

Macdonell, Arthur Anthony, 1854-1930.

*A practical Sanskrit dictionary with transliteration, accentuation, and etymological analysis throughout.*  
London : Oxford University Press, 1929. ix, 382 p. ; 30 cm.

"The original edition of this dictionary was published by Messrs. Longmans, Green & co., and has been reproduced photographically with their consent."

---

Monier-Williams, Monier, Sir, 1819-1899.

*A Sanskrit-English dictionary etymologically and philologically arranged with special reference to cognate Indo-European languages.* New ed., greatly enl. and improved. Oxford : The Clarendon press, 1899. xxxiv, [2], 1333 p. ; 31 cm.

---

## Appendix 4. Dictionaries to be Converted (cont.)

**SINDHI**

Dulamal Bulchand.

*A Sindhi-English dictionary.* Karachi, Printed at the Mercantile steam press, 1910  
[6], 2, 208, 4 p.

---

Shahani, Anandram T.

*The Sindhi-English dictionary.* 5th ed., rev. and enl. Bombay, School & College Bookstall, 1960-1969. 560 p. ; 23 cm.

---

**SINHALA**

Jayatilaka, D. B. (Don Baron), Sir, 1868-

*A dictionary of the Sinhalese language.* Colombo, The Royal Asiatic society, Ceylon branch, 1935- v. <1, pts., 1-4, 6, 8, 13-16-19; v. 20- 21; v. 23; v. 31 pt. 11 > ; 28 x 22 cm.

Vol. 1, pt. 2- compiled under the direction of Wilhelm Geiger and Helmer Smith. Vol. 31, pt. 11 has title: Simhala sabdakosaya. Vol. 1, 18; 20-21 have title: A dictionary of the Sinhala language. Vol. 1, pt. 16-17 have imprint: Ministry of Cultural Affairs, Colombo. "Begun by the Royal Asiatic Society, Ceylon Branch, and continued by the University of Ceylon, and now under the Ministry of Cultural Affairs."

---

**TAMIL**

Fabricius, Johann Philipp, 1711-1791.

*J. P. Fabricius's Tamil and English dictionary.* 4th ed., rev. and enl. Tranquebar, Evangelical Lutheran Mission Pub. House, 1972. viii, 910 p. 26 cm.

First ed., 1770, has title: A Malabar and English dictionary.

---

University of Madras.

*Tamil lexicon.* [Madras], University of Madras, 1936-6 v. 26 x 20 cm.

---

**TELUGU**

Brown, Charles Philip, 1798-1884.

*Braunya Telugu nighantu, Telugu English dictionary.* 1966. xxi, 1330 p. ; 26 cm.

Accompanied by: Dictionary of mixed Telugu: also an explanation of the Telugu alphabet.

4,791,300 bytes

---

Gwynn, J. P. L. (John Peter Lucius), 1916-

*A Telugu-English dictionary.* Delhi ; New York : Oxford University Press, 1991. xx, 569 p. ; 25 cm.

---

*Telugu-English dictionary : with the Telugu words printed in the Roman as well as in the Telugu character ...* Madras : Public Instruction Press, 1862. 484 p. ; 19 cm.

---

Sankaranarayana, Paluri.

*A Telugu-English dictionary.* Rev. and enl. ed. [with supplement]. Madras, V. Ramaswamy Sastrulu, 1964. vii, 1372 p. : port. ; 22 cm.

---

**COMPARATIVE**

Turner, R. Lister.

*A comparative dictionary of Indo-Aryan languages.* London, Oxford University Press, 1962- v. ; 29 cm. [Include supplements.]

## Appendix 5. Specimen ARTFL Dictionary Web Site

The ARTFL Project Roget's Thesaurus was selected as one of PC Magazine's Top 100 Web Sites, is one of GNN's Whole Internet Catalog Select collection, and a Magellan 3 Star Site. The Web interface is located at <[http://humanities.uchicago.edu/forms\\_unrest/ROGET.html](http://humanities.uchicago.edu/forms_unrest/ROGET.html)>.

### ARTFL Project: ROGET'S Thesaurus Search Form

**New!!** Robert Parks and the ARTFL Project are pleased to release our initial implementation of the [Wordsmyth English Dictionary-Thesaurus](#).

**New!!** The ARTFL Project and MICRA, Inc. are developing [Webster's Revised Unabridged Dictionary](#) (G & C. Merriam Co., 1913, edited by Noah Porter).

---

Search the head words or full text of *Roget's Thesaurus* version 1.02 (supplemented: July 1991) released to the public domain by [MICRA, Inc.](#) and the [Gutenberg Project](#). Internal cross references are represented as clickable hyper-text links. **Experimental**, please report problems or errors to [mark@guide.uchicago.edu](mailto:mark@guide.uchicago.edu).

---

Search the **full text** for:  (ex. threat **OR** threat\*)

To submit the query, press

Full text searches support only literal word searches and the asterisk (\*) to match word endings.

---

Search **head words** for:  (ex. threat **OR** mean. \*)

To submit the query, press

List of [head words](#) in Roget's *Thesaurus* with clickable link to corresponding entries (90 KB). Note: only head word searches support full UNIX regular expressions.

---

**HELP!** In order to develop this thesaurus together with the [Webster's dictionary](#) into an encyclopedic knowledge base for use in intelligent programs, volunteers are needed with all levels of expertise who can contribute some time. If you are willing to share your knowledge with the world, [click here for more info](#).

---

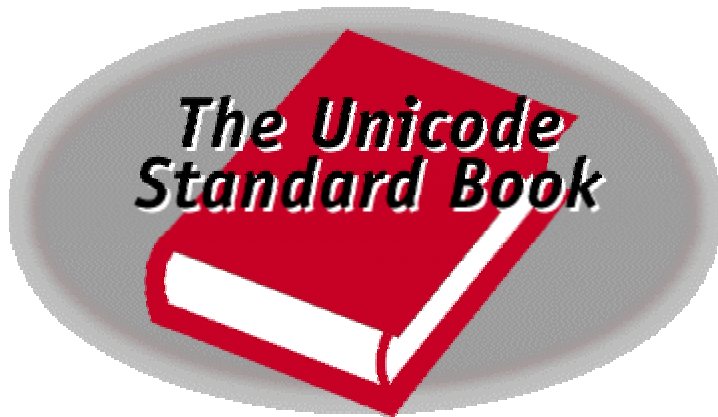
The ARTFL Project: ROGET'S Thesaurus wins [awards](#).

---

*Mark Olsen, [ARTFL Project](#), University of Chicago, [mark@guide.uchicago.edu](mailto:mark@guide.uchicago.edu)*

## Appendix 6. Statements on Digital Standards

### The Unicode Standard, Version 2.0<sup>16</sup>



*The Unicode Standard, Version 2.0*, is the authoritative source of information on the Unicode character-encoding standard, a 16-bit international character code for information processing. The rapid growth of the Internet is pushing the demand for software to be simultaneously internationalized and localized. Encompassing the principal scripts of the world, the Unicode Standard provides THE foundation for the internationalization and localization of software. Java, Windows NT, AIX,

NetWare 4.0, QuickDraw GX, are but a few of the applications, which are Unicode Standard compliant. *Version 2.0* of the Unicode Standard introduces additional scripts, and contains ten years' worth of cumulated experience from unrivaled experts in multilingual applications.

In this authorized description and guide to the Unicode Standard, you'll find documentation of all the essential aspects, including basic principles, and code charts. There is also a discussion of implementation issues and rules for Unicode Standard conformance. Scripts covered include the alphabets used in countries across Europe, Africa, Asia, and the Indian subcontinent, and the "unified Han" set of logographic characters for Chinese, Japanese, Korean and historical Vietnamese. The unified Han section includes a radical-stroke index, a multi-glyph table, and a character cross-reference chart providing mapping information to major national, bibliographic, and industrial standards. As a subset of International Standard ISO/IEC 10646-1;1993, the Unicode Standard is code value for code value equivalent with ISO/IEC 10646.

*The Unicode Standard, Version 2.0* presents standard algorithms:

- for formatting Arabic, Devanagari, and Tamil scripts;
- for display of bi-directional text (e.g., mixed English and Arabic); and,
- for the creation of Korean Hangul syllables.

Packaged with *The Unicode Standard, Version 2.0* is a CD-ROM with tables of character properties, and mappings to international, national, and vendor character sets.

*The Unicode Standard, Version 2.0* is an essential reference for computer programmers and software developers working on global software and multilingual applications. In addition, Unicode Technical Report #8: The Unicode Standard, Version 2.1 is available on this web site and provides updates and additions.

---

<sup>16</sup>Text of the description of *The Unicode Standard, Version 2.0* is reproduced from <http://www.unicode.org/unicode/uni2book/u2.html>.

## **The Dublin Core: A Simple Content Description Model for Electronic Resources**<sup>17</sup>

### **Metadata for Electronic Resources**

The Dublin Core is a metadata element set intended to facilitate discovery of electronic resources. Originally conceived for author-generated description of Web resources, it has attracted the attention of formal resource description communities such as museums, libraries, government agencies, and commercial organizations.

The Dublin Core Workshop Series has gathered experts from the library world, the networking and digital library research communities, and a variety of content specialties in a series of invitational workshops. The building of an interdisciplinary, international consensus around a core element set is the central feature of the Dublin Core. The progress represents the emergent wisdom and collective experience of many stakeholders in the resource description arena. An open mailing list supports ongoing work.

The characteristics of the Dublin Core that distinguish it as a prominent candidate for description of electronic resources fall into several categories:

#### **Simplicity**

The Dublin Core is intended to be usable by non-catalogers as well as resource description specialists. Most of the elements have a commonly understood semantics of roughly the complexity of a library catalog card.

#### **Semantic Interoperability**

In the Internet Commons, disparate description models interfere with the ability to search across discipline boundaries. Promoting a commonly understood set of descriptors that helps to unify other data content standards increases the possibility of semantic interoperability across disciplines.

#### **International Consensus**

Recognition of the international scope of resource discovery on the Web is critical to the development of effective discovery infrastructure. The Dublin Core benefits from active participation and promotion in some 20 countries in North America, Europe, Australia, and Asia.

#### **Extensibility**

The Dublin Core provides an economical alternative to more elaborate description models such as the full MARC cataloging of the library world. Additionally, it includes sufficient flexibility and extensibility to encode the structure and more elaborate semantics inherent in richer description standards.

#### **Metadata Modularity on the Web**

The diversity of metadata needs on the Web requires an infrastructure that supports the coexistence of complementary, independently maintained metadata packages. The World Wide Web Consortium (W3C) has begun implementing an architecture for metadata for the Web. The Resource Description Framework, or RDF, is designed to support the many different metadata needs of vendors and information providers. Representatives of the Dublin Core effort are actively involved in the development of this architecture, bringing the digital library perspective to bear on this important component of the Web infrastructure.

---

<sup>17</sup>Text of the statement on Dublin Core is reproduced from <<http://purl.org/DC/>>.

### Standard Generalized Markup Language and the Text Encoding Initiative<sup>18</sup>

Standard Generalized Markup Language (SGML) is an international standard for the encoding of electronic texts. The system permits interpretations of texts to be represented explicitly and without dependence on specific computing devices or programs. For example, the title on the title page of a book is usually marked by typographic conventions such as larger type size or weight of type. In SGML, that same information is noted explicitly through the use of tags (labels within angle brackets.) An opening tag or tags usually precedes text and a closing tag follows it. For example, the title of Shakespeare's play on a title page as "**Macbeth**" would be encoded "<bold><title>Macbeth</title></bold>". These tags make explicit the structure of a text and permit a reader to limit searches to particular features. Printed dictionaries are particularly rich with structure. When they are converted to electronic form, it is important to have the implicit structure made explicit for optimal computer use. This tagging and structuring gives readers the option of searching for elements within a dictionary entry such as part of speech, etymology, definition of a word, pronunciation, and translation into another language.

Text Encoding Initiative described below with an extract from their Web site, is an application of the more general SGML. As special sub-group worked to prepare detailed Text Encoding Initiative specifications for dictionaries. The full text of the base tag set for printed dictionaries they created is available at <<http://www.hti.umich.edu/bin/tei-search-idx?type=HTML&rgn=DIV1&byte=1014434>>.



#### What is the TEI?

The Text Encoding Initiative (TEI) is an international project to develop guidelines for the preparation and interchange of electronic texts for scholarly research, and to satisfy a broad range of uses by the language industries more generally. The TEI is sponsored by the Association for Computers and the Humanities (ACH), the Association for Computational Linguistics (ACL), and the Association for Literary and Linguistic Computing (ALLC). Major support for the project has come from the U.S. National Endowment for the Humanities (NEH), Directorate XIII of the Commission of the European Communities (CEC/DG-XIII), the Andrew W. Mellon Foundation, and the Social Science and Humanities Research Council of Canada.

<sup>18</sup>For a more detailed introduction to Standard Generalized Markup Language please see the Text Encoding Initiative's excellent "Gentle Introduction to SGML" at <<http://www-tei.uic.edu/orgs/tei/sgml/teip3sg/index.html>>.

## Appendix 7. ICOLC "Guidelines for Statistical Measures of Usage"

### Guidelines for Statistical Measures of Usage of Web-Based Indexed, Abstracted, and Full Text Resources<sup>19</sup>

by the International Coalition of Library Consortia (ICOLC)  
(November 1998)

#### INTRODUCTION

The use of licensed electronic information resources will continue to expand and in some cases become the sole or dominant means of access to content. The electronic environment, as manifested by the World Wide Web, provides an opportunity to improve the measurement of the use of these resources. In the electronic arena we can more accurately determine which information is being accessed and used. Without violating any issues of privacy or confidentiality we can dramatically enhance our understanding of information use.

The participating consortia of the ICOLC have a responsibility to their library members to ensure the provision of usage information of licensed electronic resources. Information providers should want the same information to better understand the market for their services as well as to create an informed customer base. These mutual interests can be best met by defining and creating a common set of basic use information requirements that are an integral and necessary part of any electronic product offering. These requirements apply to vendor operated web sites and to software provided to libraries or consortia for local operation. Information providers are encouraged to go beyond these minimal requirements as appropriate for their specific electronic resources.

These ICOLC guidelines draw heavily upon the guidelines developed by the JSTOR Web Statistics Task Force: David Farrell, Berkeley, Chair; Jim Mullins, Villanova; Kimberly Parker, Yale; Dave Perkins, CSU-Northridge; Sue Phillips, Texas; Camille Wanat, Berkeley; Kristen Garlock, JSTOR, ex-officio. The ICOLC guidelines reflect modifications to maximize their broad applicability to the diversity of resources licensed by many ICOLC members.

#### 1. REQUIREMENTS

Each use element defined below should be able to be delineated by the following subdivisions;

1. By each specific database of the provider
2. By each institutionally-defined set of IP addresses / locators to subnet level
3. By total consortium
4. By special data element passed by subscriber( e.g., account or ID number)
5. By time period. Vendor's system should minimally report by month. For each month, each type of use should be reported by hour of the day, and vendor should maintain 24 months of historical data

Use Elements that must be provided are:

Number of queries (Searches) categorized as appropriate for the vendor's information. A search is intended to represent a unique intellectual inquiry. Typically a search is recorded each time a search form is sent/submitted to the server. Subsequent activities to review or browse among the records retrieved or the process of isolating the correct single item desired do not represent additional searches, unless the parameter(s) defining the retrieval set is modified through resubmission of the search form, a combination of previous search sets, or some other similar technique.

Number of Menu Selections categorized as appropriate to the vendor's system. If display of data is accomplished by browsing (use of menus), this measure must be provided (e.g. an electronic journal site provides alphabetic and subject-based menu options in addition to a search form. The number of searches and the number of alphabetic and subject menu selections should be tracked).

---

<sup>19</sup>Text of the Guidelines is reproduced from <<http://www.library.yale.edu/consortia>>.

## Appendix 7. ICOLC "Guidelines for Statistical Measures of Usage" (cont.)

Number of sessions (Logins), if relevant, must be provided as a measure of simultaneous use. It is not a substitute for either query or menu selection counts.

Number of turn-aways, if relevant, as a contract limit (e.g., requests exceed simultaneous user limit).

Number of items examined (i.e., viewed, marked or selected, downloaded, emailed, printed) to the extent these can be recorded and controlled by the server rather than the browser:

1. Citations displayed (for A&I databases)
2. Full text displayed broken down by title, ISSN with title listed, or other title identifier as appropriate
  1. Tables of Contents displayed
  2. Abstracts displayed
  3. Articles or essays, poems, chapters, etc., as appropriate, viewed (e.g., ASCII or HTML) or downloaded (e.g. PDF, email)
  4. Other (e.g., image / AV files, ads, reviews, etc., as appropriate)

The ICOLC is preparing a separate guideline on Technical Performance of Web-based Services for reporting of system related parameters ( e.g., downtime, response time).

2. **PRIVACY AND USER CONFIDENTIALITY:** Statistical reports or data that reveal confidential information about individual users must not be released or sold by information providers without permission of the consortium and its member libraries.

3. **INSTITUTIONAL OR CONSORTIAL CONFIDENTIALITY:** Providers do not have the right to release or sell statistical usage information about specific institutions or the consortium without permission, except to the consortium administrators and member libraries. Use of institutional or consortium data as part of an aggregate grouping of similar institutions for purposes of comparison does not require prior permission as long as specific institutions or consortia are not identifiable. When required by contractual agreements, information providers may furnish institutional use data to the content publishers.

4. **COMPARATIVE STATISTICS:** Information providers should provide comparative statistics that give consortia a context in which to analyze statistics at the aggregate institutional (consortium member) level. For example, a grouping for purposes of comparison should be compiled by the information provider (e.g., statistics from an anonymous selection of similar institutions), or it might be a grouping composed on demand (e.g., statistics from all campuses in a consortium, presented either anonymously or not, as desired by the participating institutions).

5. **ACCESS / DELIVERY MECHANISMS / REPORT FORMATS:** Access to statistical reports should be provided via web-based reporting systems and be restricted by IP address or another form of security such as passwords. Institutions should be able to authorize access to their use data by other institutions in the consortium if they desire.

Information providers should maintain access to tabular statistical data through their web site (updated monthly) which a participant can access, aggregate and manipulate on demand. When appropriate, these data also should be available in flat files containing specified data elements that can be downloaded and manipulated locally. Information providers are also encouraged to present data as graphs and charts.

## Appendix 8. Impact of Technology on South Asian Studies

An extract on the impact of technology on South Asian studies from:

### *Scholarship, Research Libraries, and Global Publishing*<sup>20</sup>

#### **Technology**

The impact of technology on South Asian studies has not been as great as on area studies for some other regions. Yet, the changes are pronounced and the prospects for increased availability of scholarly resources in electronic form are significant.

The traditional sources of information related to South Asian materials are not yet converted to electronic form. Few libraries in South Asia have their catalogs converted to machine-readable form. Retrospective conversion of U.S. and South Asian library catalogs remains an important desideratum for scholarly access and a prerequisite for cooperative collection development. The Committee on South Asian Libraries and Documentation (CONSALD) has a plan for retrospective conversion, but progress remains difficult.

The picture is the same for bibliographies and reference tools. The world's major online indexes and abstracts cover relatively few publications from South Asia. Further, most of the standard South Asia bibliographies and reference tools are only available as printed tools. With the exception of journal indexing projects noted above, not even the most basic reference sources on South Asia are accessible electronically.

Several ARL [Association of Research Libraries] libraries are taking advantage of advances in client/server technology. Integrated sites, the South Asia Gopher at Columbia University, the Asian Studies Network Information Center at the University of Texas, and the Web pages at several institutions are serving important functions in providing a structured home for existing electronic resources and in encouraging the production of new resources. But humanities and social sciences libraries in South Asia have lagged far behind science libraries in use of electronic technology, largely for lack of financial support.

There is, to date, no systematic planning to expand the availability of full-text databases or scanned images of photographs and rare texts for South Asian studies. Faster and more reliable communications – use of electronic mail, list servers, and the World Wide Web – may draw more scholarly and library colleagues from South Asia into the mainstream of inquiry that originates outside the region. Yet lack of funding for creating electronic resources remains a formidable obstacle.

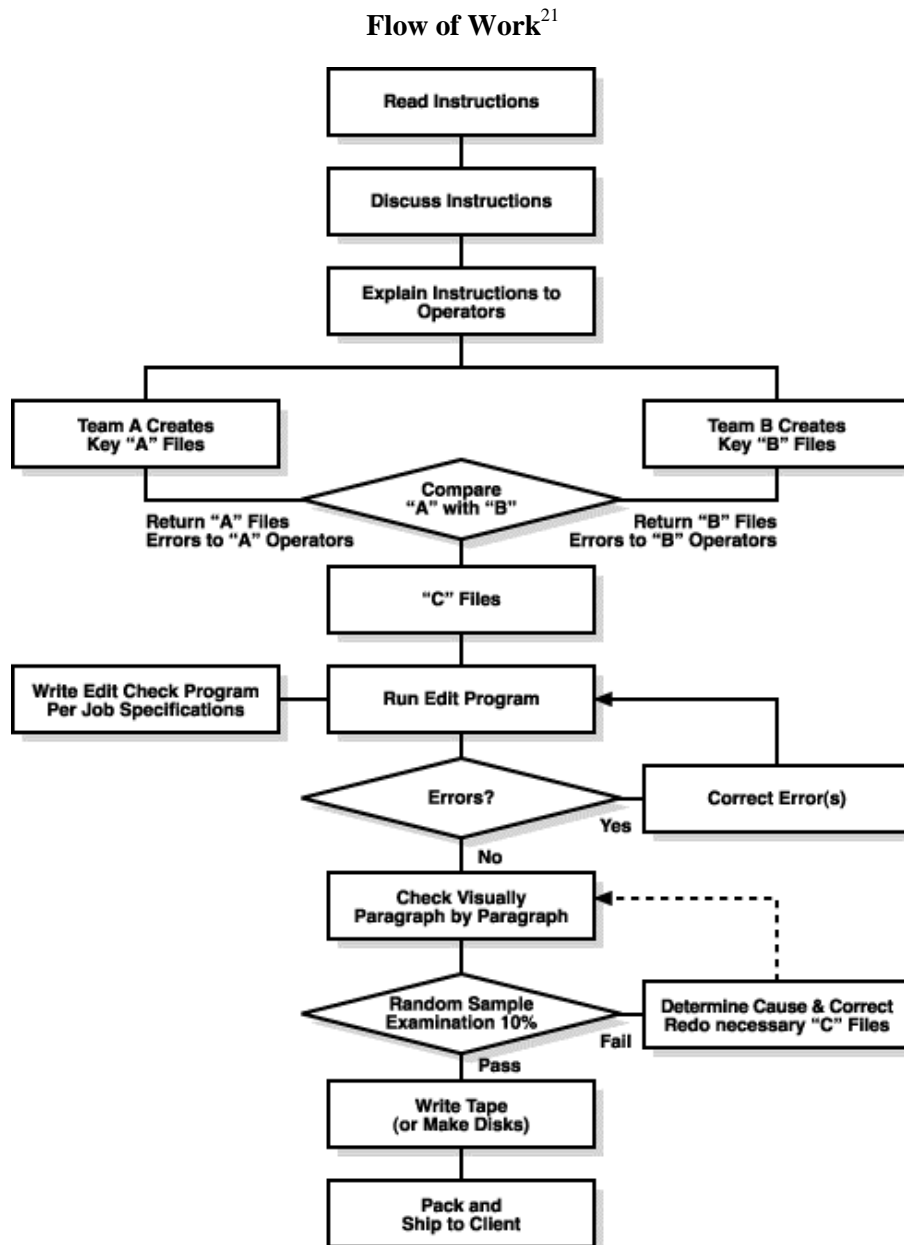
In conclusion, one notes the challenges presented to research libraries by the expanding range of scholarly inquiry about the South Asian subcontinent and an increasing move toward inter-disciplinary studies. These changes in research have resulted in calls for more facile access to a wider array of resources, both in print and in electronic form. Considering all these elements, the need for new sources of funding and closer collaboration among North American libraries and also with counterparts elsewhere in the world are major requirements if libraries on South Asia are to justly meet the needs of researchers.

---

<sup>20</sup>Jutta Reed-Scott, *Scholarship, research libraries, and global publishing: the result of a study funded by the Andrew W. Mellon Foundation* (Washington, DC: Association of Research Libraries, 1996), p. 82-83.

**Appendix 9. Data Entry by Double Keying**

The Input Center, a Chicago firm, offers high quality data entry work with 99.99% accuracy rate by double keying, comparing, and proofreading each job. This is much greater accuracy than is possible with OCR (Optical Character Recognition) scanning which usually only achieves 95% accuracy. OCR scanning would produce 50 errors every 1,000 characters. Double keying produces only 1 error every 10,000 characters.



<sup>21</sup>The flow chart is copied from <<http://www.inputcenter.com/PAGES/SERVICES/PROCESS.HTM>>.

## **Appendix 10. Response to GEPA Section 427**

### **The University of Chicago's commitment to increasing the number of minority group members, women, and students with disabilities**

In keeping with its long-standing traditions and policies, the University of Chicago, in admissions and access to programs, considers students on the basis of individual merit and without regard to race, color, religion, sex, sexual orientation, national or ethnic origin, age, disability, or other factors irrelevant to participation in the programs of the University.

The University of Chicago has a long record of searching widely for excellent students and committing support to these students once they matriculate. Special efforts are made to increase the number of minority and women students. Of late years, in addition, special efforts have been made all over the University to accommodate students with disabilities. This is a University with a remarkably diversified student population.

The University puts responsibility for implementing such policies and practices into many hands, as it should: department chairmen, deans, deans of students, admissions officers, supervisors, department heads, human resources management staff, etc. A central committee, "The Provost's Committee on Minority Issues," coordinates the work of the University in connection with the many aspects of this work.

Looking at the specific efforts in the whole University: Minority, women, and disabled students benefit from the special undergraduate recruiting program, "Setting Sights;" the summer program of research opportunities for minority group members and women, the Summer Research Opportunity Program (SROP), conducted in coordination with the Consortium for Institutional Cooperation (CIC); the Resource Guide and newsletter published to assist minority students; the counseling and tutoring programs which focus on individualized assistance for both minority students and students with disabilities; our selection in 1992 as a host school for the Consortium On Financing Higher Education (COFHE) Pipeline Project; and assistance for students in applying for United States Department of Education Patricia Roberts Harris fellowships. The University and Chicago State University have a continuing collaborative effort for undergraduate and graduate students, supported under a grant to the University from the Illinois Board of Higher Education,

The University believes that in addition to campus efforts, the University must have community outreach programs. For that reason, the University has many projects in local public schools, an outreach program in the Argonne Laboratory for high school students interested in science, and numerous students involved in tutoring programs in local institutions. Of special note is the Office of Special Programs, a program located in the University for more than twenty years, that prepares inner-city youngsters for college admission and graduation through its Pilot Enrichment Program (PEP) and its Upward Bound Program, its Blum-Kovler Teacher Awards and in cooperation with the Institute of Athletics and Education.

The University's commitment has been steadfast and the progress has been forward.

### **The University of Chicago's commitment to removing architectural barriers**

The University of Chicago has a long record of working successfully to provide access to individuals with disabilities. Over the years, many students, faculty, and staff with disabilities have studied, completed degrees, and worked at the University; and many visitors have given and attended public lectures, exhibits, functions, and the like.

Most recently in 1991, the University examined its approximately ninety buildings to determine which buildings were accessible and which needed modification. Since that study was completed in 1992, the University has undertaken several renovation projects each year with, in most cases, the sole purpose of removing architectural barriers. These projects have included installation of numerous ramps, automatic door openers, wheelchair lifts, elevators, curb cuts, handrails, and signage; modification of rest rooms (including shower stalls in the athletic facilities) and elevator tabs; lowering of public telephones

## Appendix 10. Response to GEPA Section 427 (cont.)

and provision of amplified handsets; striping of special parking spots; and numerous other improvements to access. Of course, all new construction is compliant and accessible.

The University's commitment to inclusion and accessibility has been steadfast and progress continues.

### **University of Chicago Affirmative Action Policy**

The University of Chicago has a policy of affirmative action and of providing equal opportunity in employment for all qualified persons and prohibiting discrimination in employment on the basis of attributes unrelated to performance or the job such as race, color, religion, sex, sexual orientation, national or ethnic origin, age, disability, or veteran status. This policy includes the commitment to maintain a work environment free from harassment.

Dissemination of Policy. Statements of the University's policy are carried in University personnel policy booklets distributed to all entering employees and available to all employees at any time. Periodic reports are published in the *Record*. Posters of policy and the law are displayed on personnel bulletin boards at various points around the campus. Meetings have been held with department heads and supervisors to discuss the affirmative action program. The existence of the affirmative action program and obligations to follow it is widely known throughout the University.

The policy is also made known outside the University. Advertisements for openings carry the notice that the University is an equal employment opportunity / affirmative action employer. Disability assistance availability is noted on posters and announcements of campus events. Announcements of searches for employees, sent to various institutions and organizations, employment agencies, and the like, mention the University's equal employment opportunity / affirmative action policy. Through all these activities, prospective employees are made aware of the University's commitment to this program.

Responsibility for Implementation. The responsibility for implementing the affirmative action program is recognized at the highest levels of the University's administration. The actual performance of affirmative action is a daily activity that takes place within each department and other administrative units of the University. Members of the faculty and staff have been made aware of the University's strong commitment to affirmative action. In the academic departments, it is the responsibility of each departmental faculty and its chairman in recommending academic appointments to assure proper compliance with the policy. The deans of the academic divisions have the responsibility to review the activities of the departments in appointments, reappointments, promotions, and other personnel actions relating to faculty and other academic appointments to assure that the affirmative action program is followed. The Provost, the chief academic officer of the University, in reviewing and approving actions such as recommendations for appointments sent forward from the deans, considers whether the appointments are consistent with all University policies, including its affirmative action policy. Should he/she perceive a problem to exist, he/she exercises his or her authority to take necessary action. In addition, at every level of the process, the faculty appointment committees, chairmen, deans, and Provost may consult with the Affirmative Action Officer about the interpretation of the program and its application to particular situations. Such consultations take place often.

The Affirmative Action Officer has a broad spectrum of duties in connection with equal opportunity and affirmative action. Under the general supervision of the President, and in regular consultation with the Provost and Vice-President for Administration, the Affirmative Action Officer is concerned with the ongoing process of developing and supplementing the affirmative action program and disseminating information about the program within and outside the University. In the development of the affirmative action program, the Affirmative Action Officer participates in identifying problems and their solutions and in overseeing the collection of data, the determining of availability, and the making of realistic hiring projections. The Affirmative Action Officer sees that there is adequate review of the variety of actions involved in affirmative action, including recruitment, hiring, compensation administration, reappointment and promotion, training programs, posting of nondiscrimination notices, providing of facilities on a nondiscriminatory basis, providing guidance to supervisors, investigating and resolving informal complaints, and supervising responses to enforcement agencies.