U.S. Department of Education

Washington, D.C. 20202-5335

APPLICATION FOR GRANTS UNDER THE

INTERNATIONAL RESEARCH AND STUDIES PROGRAM CFDA 84.017A. SCHEDULE 84.017A-1
CFDA # 84.017A
PR/Award # P017A090335
Grants.gov Tracking#: GRANT10266502

OMB No. 1840-0795, Expiration Date: 08/31/2010
Closing Date: APR 23, 2009

Table of Contents

Forms

1. *Application for Federal Assistance (SF-424)* e1
2. *Standard Budget Sheet (ED 524)* e5
3. *SF 424B - Assurances Non-Construction Programs* e7
4. *Disclosure of Lobbying Activities* e9
5. *427 GEPA* e10
   - *Attachment - 1* e11
6. *ED 80-0013 Certification* e12
7. *Dept of Education Supplemental Information for SF-424* e13

Narratives

1. *Project Narrative - (Abstract Narrative...)* e14
   - *Attachment - 1* e15
2. *Project Narrative - (Project Narrative...)* e16
   - *Attachment - 1* e17
3. *Project Narrative - (Other Narrative...)* e48
   - *Attachment - 1* e49
4. *Budget Narrative - (Budget Narrative...)* e73
   - *Attachment - 1* e74

This application was generated using the PDF functionality. The PDF functionality automatically numbers the pages in this application. Some pages/sections of this application may contain 2 sets of page numbers, one set created by the applicant and the other set created by e-Application's PDF functionality. Page numbers created by the e-Application PDF functionality will be preceded by the letter e (for example, e1, e2, e3, etc.).

U.S. DEPARTMENT OF EDUCATION

BUDGET INFORMATION

NON-CONSTRUCTION PROGRAMS

OMB Control Number: 1890-0004

Expiration Date: 06/30/2005

Name of Institution/Organization:
CRCL INC

Applicants requesting funding for only one year should complete the column under "Project Year 1." Applicants requesting funding for multi-year grants should complete all applicable columns. Please read all instructions before completing form.

SECTION A - BUDGET SUMMARY

U.S. DEPARTMENT OF EDUCATION FUNDS

Budget Categories	Project Year 1(a)	Project Year 2 (b)	Project Year 3 (c)	Project Year 4 (d)	Project Year 5 (e)	Total (f)
1. Personnel	$ 103,250	$ 107,380	$ 111,675	$ 0	$ 0	$ 322,305
2. Fringe Benefits	$ 30,476	$ 31,695	$ 32,963	$ 0	$ 0	$ 95,134
3. Travel	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
4. Equipment	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
5. Supplies	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
6. Contractual	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
7. Construction	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
8. Other	$ 9,500	$ 22,500	$ 22,500	$ 0	$ 0	$ 54,500
9. Total Direct Costs (lines 1-8)	$ 143,226	$ 161,575	$ 167,138	$ 0	$ 0	$ 471,939
10. Indirect Costs*	$ 11,458	$ 12,926	$ 13,371	$ 0	$ 0	$ 37,755
11. Training Stipends	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
12. Total Costs (lines 9-11)	$ 154,684	$ 174,501	$ 180,509	$ 0	$ 0	$ 509,694

***Indirect Cost Information *(To Be Completed by Your Business Office)*:**

If you are requesting reimbursement for indirect costs on line 10, please answer the following questions:

(1) Do you have an Indirect Cost Rate Agreement approved by the Federal government? [ ] Yes [X] No
(2) If yes, please provide the following information:
Period Covered by the Indirect Cost Rate Agreement: From: __/__/____ To: __/__/____ (mm/dd/yyyy)
Approving Federal agency: [ ] ED [ ] Other (please specify): _____
(3) For Restricted Rate Programs (check one) -- Are you using a restricted indirect cost rate that:
[ ] Is included in your approved Indirect Cost Rate Agreement? or, [ ] Complies with 34 CFR 76.564(c)(2)?

ED Form No. 524

U.S. DEPARTMENT OF EDUCATION BUDGET INFORMATION NON-CONSTRUCTION PROGRAMS	OMB Control Number: 1890-0004 Expiration Date: 06/30/2005
Name of Institution/Organization: CRCL INC	Applicants requesting funding for only one year should complete the column under "Project Year 1." Applicants requesting funding for multi-year grants should complete all applicable columns. Please read all instructions before completing form.

SECTION B - BUDGET SUMMARY
NON-FEDERAL FUNDS

Budget Categories	Project Year 1(a)	Project Year 2 (b)	Project Year 3 (c)	Project Year 4 (d)	Project Year 5 (e)	Total (f)
1. Personnel	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
2. Fringe Benefits	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
3. Travel	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
4. Equipment	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
5. Supplies	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
6. Contractual	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
7. Construction	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
8. Other	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
9. Total Direct Costs (lines 1-8)	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
10. Indirect Costs	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
11. Training Stipends	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
12. Total Costs (lines 9-11)	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0

Project Narrative

Abstract Narrative

Attachment 1:
Title: Pages: Uploaded File: **1235-irs-abstract.pdf**

No More Data Silos – Interoperable Digital Resources in Title VI

In 2006, a committee to review Title VI and Fulbright-Hays *International Education Programs* was established by the National Research Council at the request of Congress and the Department of Education. Its final report, *International Education and Foreign Languages: Keys to Securing America's Future*, was published by the National Academy of Sciences in 2007.

"Advancing uses of new technology in foreign language and international studies" was a key Congressional concern. The NRC committee reported that current efforts to "*effectively apply developments in technology to … support of language instruction suffer from a dispersion of resources*," and that TitleVI/FH programs "*could do more to maximize the potential created by current technologies*." They stressed the need for "*generic tools*" and "*common platforms*" in using technology, and concluded:

> "*A national technology infrastructure could significantly enhance the nation's capacity for education in critical and less commonly taught languages"*

Traditional approaches have not met this challenge. Harvesting metadata and building databases help inventory dispersed content, but they don't unite standalone *data silos*. Haystacks are easier to find, yet needles – e.g. definitions of words in less-common languages – remain elusive.

But helping teachers build common platforms (or programmers write generic tools) that query *all* Title VI projects for entries in *every* known dictionary, audio, or image resource could be as easy as embedding Google maps or Amazon book searches – if we, like them, had a **technology infrastructure**, built from **interoperable resources**, and delivered using Web-based **services**. This project will do the basic research and surveys required to get us there. We will:

- Survey and identify specialized materials already developed with Title VI funding, and appropriate for adaptation or implementation as **digital lexical resources**.
- Research and develop "best practices" guidelines for providing access to these resources via **interoperable Web-based services.**
- Test, evaluate, and promulgate these "best practices" by building a **reference model** of **generic tools** and **common platforms**, including a one-click dictionary search tool.

These tangible results will actively involve current stakeholders in Title VI programs, and help assemble the critical mass that brings **voluntary adherence** to "best practice" guidelines.

Our primary focus is on **bilingual dictionaries** in electronic and paper form, because of:

- their central importance to foreign language study and research at all levels, including K-12 and postsecondary education,
- the opportunity they provide to address this IRS competition's **competitive priority** on a broad scale, dealing with the many critical LCTLs that use non-Roman scripts,
- their suitability to task as **cornerstone Web services** – such as dictionary lookup – that will enable and encourage leading-edge infrastructure applications, and
- the huge existing (and ongoing) investment made by the Title VI programs – in just the past five years, 17 IRS and TICFIA-funded projects have created resources for Albanian, Arabic, Cambodian, Hindi, Nepali, Swahili, Tamil, Uzbek, Yoruba, and dozens more,

The proposal is submitted by the *Center for Research in Computational Linguistics*, which pursues a highly successful program of lexical resource development with NEH, NSEP, and US/ED support. We are joined by an advisory board of federal grantees, including IRS and TICFIA project directors who will collaborate in making project data and services available.

Project Narrative

Project Narrative

Attachment 1:
Title: Pages: Uploaded File: **1238-irs-narrative.pdf**

No More Data Silos – Interoperable Digital Resources in Title VI

1. NEED FOR THE PROJECT 1
1.1 INTRODUCTION 1
1.2 RECOGNIZING THE PROBLEM 2
1.3 BUILDING CAPACITY TO SOLVE THE PROBLEM 3
2. USEFULNESS OF EXPECTED RESULTS 5
3. DEVELOPMENT OF NEW KNOWLEDGE 9
4. FORMULATION OF THE PROBLEMS AND KNOWLEDGE OF RELATED RESEARCH (10) 11
4.1 FORMULATION OF THE PROBLEM 11
4.2 KNOWLEDGE OF RELATED RESEARCH 12
5. SPECIFICITY OF STATEMENT OF PROCEDURES 14
6. ADEQUACY OF RESEARCH METHODOLOGY AND SCOPE OF PROJECT 16
6.1 ADEQUACY AND PROPRIETY OF RESEARCH AND TESTING 16
6.2 SIZE, SCOPE, AND DURATION OF THE PROJECT 16
7. PLAN OF OPERATION 18
7.1 DESIGN 18
7.2 MANAGEMENT AND ADMINISTRATION 18
7.3 RELEVANCE TO THE PURPOSES OF THE IRS (SECTION 605) PROGRAM 19
7.4 USE OF RESOURCES 19
7.5 EQUAL ACCESS 20
8. QUALITY OF PERSONNEL 21
8.1 PROJECT DIRECTOR 21
8.2 OTHER PERSONNEL 21
8.3 TIME COMMITMENTS 23
8.4 BROAD PARTICIPATION 23
9. BUDGET AND COST EFFECTIVENESS 24
9.1 ADEQUACY OF BUDGET 24
9.2 REASONABLENESS OF COSTS 24
10. EVALUATION PLAN 25
10.1 IDENTIFCATION OF PARTICIPANTS AND STAKEHOLDERS 26
10.2 SURVEY OF EXISTING RESOURCES AND PROJECT CAPACITY 26
10.3 STUDY OF CURRENT PRACTICES AND NEEDS FOR INTEROPERABILITY 27
10.4 REFERENCE MODEL FOR TOOLS AND PLATFORMS 28
11. ADEQUACY OF RESOURCES 29
11.1 FACILITIES 29
11.2 EQUIPMENT AND SUPPLIES 29
12. COMPETITIVE PRIORITY 30

TABLE 1 SAMPLE WEB-BASED SURVEY INSTRUMENT 15
TABLE 2 PROJECT TIMELINE AND ACHIEVEMENT MILESTONES 18
TABLE 3 RESPONSIBILITIES OF PROJECT STAFF 20
TABLE 4 ADVISORY BOARD FOR THE PROJECT 23
TABLE 5 TIME COMMITMENTS OF PROJECT PERSONNEL 23

TABLE A1 PROJECT TIMELINE A1

APPENDIX
A PROJECT TIMELINE A1
B LETTERS OF SUPPORT AND COLLABORATION A2
C CURRICULA VITAE A11

1. NEED FOR THE PROJECT

In the past five decades, the Title VI programs have supported the development or digitization of well over one hundred bilingual dictionaries. Yet, there is no simple way to look up a single word in *all* of them. We can link to project Web pages, but direct access to (and reuse of) project *data* remains tantalizingly out of reach.

This proposal addresses the specific problem of enabling such searches among existing Title VI resources, and the general problem of guiding future resource development down the path of *interoperability*. In doing so, we speak to the **present** and **future needs** of Title VI, as articulated in a 2007 National Research Council program review (henceforth *NRC*),[1] passed into law in the 2008 **Higher Education Act Reauthorization**, and discussed below. Our proposed research can be summed up in four words: *resources*, *capacity*, *services*, and *protocols*.

- What **resources** have Title VI projects created? What is their condition/copyright status?
- What **capacity** do Title VI projects have to provide data and services?
- What **services** (e.g. what kinds of queries) should be supported?
- What **protocols** should be recommended for providing those services?

Below, Section **1.1** briefly describes the project, Section **1.2** documents the expressed need that prompts it, and Section **1.3** outlines the steps we will take to solve the underlying problems.

1.1 INTRODUCTION

This project addresses issues central to current and future development of **digital resources** for foreign-language education and research, including Web-based *dictionaries*, text and bitext *corpora*, and audio and image *databases*. Focusing on **bilingual dictionaries**, we will:

- survey and identify specialized materials already developed with Title VI funding, and appropriate for adaptation or implementation as **digital lexical resources**.
- research and develop "best practices" guidelines for providing access to these resources via **interoperable Web-based services**, and

[1] *International Education and Foreign Languages: Keys to Securing America's Future* (2007), Committee to Review Title VI and Fulbright-Hays International Education Programs, O'Connell, M.E. and Norwood, J.L., eds., National Research Council, National Academies Press.

- test and evaluate the effectiveness of these "best practices" by creating and distributing a **reference implementation** of basic generic tools and common platforms.

In short, we will create a road map for digital resource development and sharing in Title VI. We will begin by developing a detailed picture of the resources currently at our disposal. Next, we will address technical aspects of access and interoperability, and recommend "best practices" for planning future projects and retrofitting past efforts. Finally, we will work with collaborators from both ongoing and completed IRS, TICFIA, and other US/ED-funded projects to test and demonstrate our recommendations: implementing, and building a reference application that is based on a set of interoperable digital dictionary resources.

The project is expected to have a significant impact on guidelines for, and design of, resource-development proposals in IEPS initiatives. Most funded efforts are planned by linguists or educators with little software experience, and no program guidance. They yield isolated data silos – not the "common platforms" the NRC review of Title VI calls for, and not the **systematic dissemination of data** the Higher Education Act demands. We will help project planners – and Title VI policy makers – understand and work toward the shared goal of interoperability.

The project will also encourage and enable the development of proposals to build generic tools that make use of Title VI resources, linking the data silos together. To our knowledge, such proposals have *never* been made because no two data development projects have *ever* conformed to the same standards, or to a recognized set of "best practices."

Prepared in accordance with the authorized activities of Section 605(a), the "best practices" we recommend will directly address both issues (see also Section **7.3**, below). Based on our research into current capacities and future needs they will **improve immediate results**, and enable and encourage proposals for general-purpose tools that **further long-term outcomes**.

1.2 RECOGNIZING THE PROBLEM

Title VI programs produce an enormous variety of instructional and assessment materials. These include text, image, audio, and visual resources, and are distributed in every possible physical (paper, CD, DVD) and electronic (FTP, Web-based) form.

Despite – or perhaps because of – this inarguable success, there is a broad perception that the program as a whole **does not equal the sum of its parts**. A 2006 call for public comment on the Title VI (and Fulbright-Hays) programs by the National Research Council drew repeated references to the same points under the rubric of "program weaknesses:"

> "The major criticism of [IRS] is that the materials and information it produces are not sufficiently disseminated or made readily available to the public." *David Wiley*

> "The following needs should be addressed: … Systematic collection and dissemination of information, materials and outcomes in the form of digitized databases." *Uliana Gabara*, Association of International Education Administrators.

> "An enormous amount of expertise and information/research is produced by Title VI and Fulbright-Hays programs, but is not easily accessible to stakeholders and the public. . … We believe what's needed is a comprehensive, digitized database involving the collection and wide dissemination of information and materials, including best practices." *William Brustein*, Coalition for International Education [*which represents 3,300 colleges and universities*].

The Council's final report accepts and extends this commentary. Chapter 8, **Technology**, ends:

> "Conclusion: current efforts to effectively apply developments in technology to … support of language instruction suffer from a dispersion of resources. … Title VI/FH programs are using available technologies, such as the Internet and distance learning, but they could do more to maximize the potential created by current technologies."

And Chapter 12, **Conclusions and Reformations**, notes that:

> "A national technology infrastructure could significantly enhance the nation's capacity for education in critical and less commonly taught languages across the federal programs designed to teach them."

Ultimately, the NRC report calls for initiating a national project whose initial focus should be:

> "research and development needed to design and implement a range of new technology-based methods for (1) assessing language proficiency and (2) supporting language instruction through the development of common platforms."

1.3 Building capacity to solve the problem

The current proposal will help to assess and develop the present and future capacity required to meet the NRC goals. We do so by identifying, and then working to correct, several serious disconnects between Title VI program aspirations and provision of the supportive measures required to achieve them. For example, the report's call for "generic tools," below, neglects a

critical connection: even digital content must follow generally accepted, documented standards if generic, language-independent development is to be feasible:

> "Given the nature of digital content, these tools can be language independent—that is, separate from the language under consideration—because the content is digital and thus can be manipulated with generic tools." (NRC)

The "best practices" we recommend will **fill this gap**.

On-line databases of digital resources are a second long-time program aspiration. Calls for these have, in fact, been heeded in the past, leading to construction of the *e-LCTL DL-Database* and others, discussed further in Section **3**.

But with rare exceptions, these databases do not actually store the results of Title VI projects. Rather, they locate, describe, and provide links to resources. Thus, they remain one step removed from the goal: ensuring that data is *preserved*, and remains *accessible*. Here, the disconnect is literal: absent standards for Web-based data access and preservation, databases are undermined when these links are cut after a few years.

Perhaps the most serious disconnect is found between long-term program goals, and the means by which individual projects are evaluated in the here and now. As the NRC report notes:

> "Program monitoring and performance measurement efforts have largely been conducted as a top-down enterprise accompanied by little consultation with grantees. Yet program performance is also partially dependent on the extent to which universities and other grantees share program goals and work collaboratively to accomplish those goals." (NRC pg. 211)

Current means of evaluating digital resources – counting website visits or DVD sales – do not adequately recognize, and can even discourage, project activities whose primary benefit targets the community and not the developer. We need to make it clear that building the national technology infrastructure by enabling **transparent access to data and services** is our shared goal, and then find means of fairly recognizing, evaluating, and rewarding this contribution.

The disconnects we cite have not arisen through any lack of care or ability. Accessibility and interoperability have not suffered because *wrong* decisions were made in project design. Rather, they have not been central design considerations at all, because the critical mass of users and resources that would benefit from "common platforms" had not yet accumulated. It has now.

2. USEFULNESS OF EXPECTED RESULTS

Broad, voluntary adherence to "best practices" for data access is our goal. The surest evidence that the project's:

> "*anticipated research results have a good potential for being utilized by other projects or programs for similar educational purposes*"

is seen in the other US/ED projects, past and present, large and small, that have chosen to collaborate and/or join our Advisory Board. They include:

- *Digital Dictionaries of South Asia* (IRS/TICFIA, 1999/2002/3/5, James Nye)
- *Tibetan and Himalayan Digital Library*. (IRS/TICFIA 2000/2/3/5, David Germano)
- *Southeast Asian Languages Library* (TICFIA 2005, Robert Bickner)
- *Nahuatl Learning Environment* (IRS 1999, Jonathan Amith)
- *Interactive Intermediate Vietnamese* (IRS 1995, Stephen O'Harrow)
- *Computer Dictionaries for Thai and Northern Indonesian* (IRS 1992, George Henry)
- *Dictionary of Cebuano Visayan* (ED, 1972, John. U. Wolff)
- *LangNet* (FIPSE 2000, Catherine Ingold)
- *Digital Dictionaries for Iraqi, Moroccan, and Syrian Arabic* (IRS 2007, M. Maamouri)
- *Slavic and East European Language Resource Center* (LRC 1999/2002, Edna Andrews)

These projects include dozens of dictionaries (and in some cases other resources) in a variety of forms, some entirely in paper, and some conceived and delivered as born-digital resources. This broadly representative group was intentionally assembled with two goals in mind:

- to confront issues relevant to the **widest possible range** of stakeholder projects, and
- to establish the **critical mass** of **early adopters** needed to ensure the long-term success of the project.

Below, we discuss the value and potential for utilization of the project's three main results.

Result 1: *Survey and identify specialized materials already developed with Title VI funding, and appropriate for adaptation or implementation as digital lexical resources.*

Useful because: We must understand both the resources and capacities of Title VI grantees.

As noted at the start of Section 1, we focus on digital dictionaries, along with associated resources like text and audio corpora. Such resources are central to the language-teaching enterprise, and are widely produced by Title VI programs.

Past development has specialized in **components** (the "data silos" referred to earlier – usually standalone websites) rather than **systems**. We now have dozens of dictionary resources, each excellent in its own way; yet broad language instruction programs (like the US/ED-funded *LangNet* program) do not make use of them. Why not? It is because they do not share common platforms: including each resource would require expensive, time-consuming, custom work.

Our surveys will address the dual questions of **resources** and **capacity**. A sample survey instrument is shown in Section **5**, below; queries include:

- *Resources***:** what is their current form? If electronic, what are the external (text, doc, database) and internal (XML-tagged, labeled, font/layout delimited) data formats? Are non-Roman scripts required? Non-Unicode font encodings? Are there copyright restrictions? Can the material be redistributed? Can it be data-mined for transformative "fair-use" applications? What kinds of queries can the content support?
- *Projects***:** who is the technical lead / contact point? Does the project manage its own website? Control its current search-engine technology? What queries can reasonably be supported? Is there interest in collaborating on "best practices" development?

These questions have not been addressed by other surveys (in particular, a current JBLA / CAL *Evaluation of the Impact of the International Research and Studies Program*, with which we are cooperating). Yet answers to them are a necessary prerequisite for developing a national resource infrastructure: **we cannot move forward until we know where we stand now**.

Result 2: *Research and develop "best practices" for providing access to resources.*

Useful because: Title VI funding programs and grantees will use these guidelines to propose and/or retrofit interoperable resources.

There is tremendous demand within the field for this assistance. In contacting Title VI project directors past and present while preparing this proposal, we have been struck by their

enthusiasm, and willingness to participate. Project directors want their results to be used as widely as possible, and more importantly, Title VI stakeholders want to use them.

The same sense emerged in the follow-up to a talk, *Digital Resources in Title VI: planning the next 50 years*, presented at the Title VI 50th Anniversary Conference.[2] There is a growing awareness that the various Title VI programs must begin to present their results more effectively; not simply in terms of statistics or evaluations, but by learning how to induce projects to make taxpayer-funded content more accessible.

Currently available alternatives do not suffice. At present, two strategies are followed:

- *Websites* / *portals to websites*: require the user to go to a site in order to pose queries; creates the very problem of data silos that we are trying to solve.
- *Databases* / *archives*: in theory, able to redistribute data sets; but in practice, almost entirely ignored by Title VI community projects that will not – or cannot – give away complete dictionaries and/or other resources.

These are not the only viable alternatives. For example, consider Google. It has a database. It has a website. But it does not require users to download the former, or to visit the latter. Rather, Google (like most modern data enterprises) provides simple ways to let developers build queries into web browsers, web pages, or mouse clicks.

We will adapt the methods of modern enterprises to the needs of the Title VI programs, and provide **access to data services** through the use of **Web APIs** (an API, or *application programming interface*, lets Internet-based programs share and access data). At present, we know of only one federal website that provides API services: **USAspending.gov**, set up to disclose federal expenditures (*The People's Data*, Feb. 28 2009, accessible at *www.newsweek.com/id/186991*).

Web APIs avoid the many difficulties that US/ED projects associate with sharing data sets: it is hard to standardize formats (because the types of data involved, from simple glossaries to complex analytical dictionaries, vary so widely), and it is impossible to mandate sharing because data is increasingly licensed – not developed – by Title VI projects. Instead:

[2] Doug Cooper, March 2009, Washington DC. Available at http://sealang.net/archives/title6-50.pdf

- rather than distributing data sets in toto, we provide ready access to their *contents*,
- rather than *isolating* data sources, we provide a means to *merge* them,
- rather than restricting data to a specific context, we let it be presented by flexible tools,
- rather than requiring fixed dataset snapshots, we let data be updated constantly.

We will not prescribe implementation standards for these services – every project has its own needs and abilities. Rather, the thrust of "best practices" is to identify a common-sense core set of necessary functional service guidelines, and to point practitioners to the best known approaches to achieving them. And, once again, answers to these questions are necessary for *any* forward movement in developing a **national resource infrastructure**.

Result 3: *Demonstrate and test the effectiveness of these "best practices" by creating and distributing a reference set of basic generic tools and common platforms.*

Useful because: Builds the critical mass that ensures the utilization and success of our results.

Best practices for interoperability cannot be selected or promulgated in the abstract. Rather, they must be demonstrated in practice, in a manner that involves primary stakeholders:

- they must be acceptable to current projects – data *developers*,
- they must be practical (and comprehensible) to tool builders – data *enhancers*,
- they must support the needs of students, teachers, and researchers – data *consumers*.

As discussed above, there is a tremendous demand for these practices to be articulated, for the perhaps surprising reason that competition for federal funding is increasingly competitive. The stakes rise constantly; not simply in the quality of work that must be produced, but in the demand for broad inter-institutional involvement, and in the level of digital-delivery mechanisms that must be provided. In this environment, "best practices" provide:

- *traction:* a proposal to build a tool that helps provide access to interoperable resources for fifty languages is more persuasive than a proposal that is only plug-compatible with one.
- *leverage:* scholars can concentrate on what they do best – such as compiling dictionaries – while taking advantage of a growing base of end-user applications: software tools that deliver dictionaries, vocabulary drills, writing exercises, and so on.

3. DEVELOPMENT OF NEW KNOWLEDGE

The proposal develops new knowledge in four areas:

- What **resources** have Title VI projects created? What is their condition, accessibility, and copyright status? Do any (e.g. LCTL lexical resources) require special technology?
- What **capacity** do Title VI projects have to provide data and services?
- What **services** (e.g. what kinds of queries) should be supported?
- What **protocols** should be recommended for providing services?

The answers to these questions, and the "best practice" guidelines they inform, directly address each paragraph in section (b), **Purposes**, of Title VI under Part A:

Paragraph 1(c): "*to develop and validate specialized materials and techniques for foreign language acquisition and fluency*"

Paragraph 2: "*to support cooperative efforts promoting access to and the dissemination of international and foreign language knowledge, teaching materials, and research … through the use of advanced technologies*"

Paragraph 3: "*to coordinate the programs of the Federal Government in the areas of foreign language, area studies, and other international studies, including professional international affairs education and research*."

Our results will have a substantial impact on materials development and usage. They will help:

- make it possible for materials to be developed or adapted to a **common specification**,
- make it possible for materials to be accessed and disseminated via **common protocols**,
- make it possible to develop new tools and techniques that are **not language specific**, and which extend the benefits and utility of **existing resources**, and
- **enable cooperation and coordination** between unrelated programs, working directly at the data level, in true "Web 2.0" fashion (see below).

At present, resources developed under Title VI programs like IRS and TICFIA are isolated from one another, with little if any cooperation between projects. Yet as obvious as the need is to "*maximize the potential created by current technologies*," we have an unexpectedly poor understanding of what resources we can currently draw on, how current technologies might be applied to them, or even what sorts of "potential" might be realized.

What is clear, however, is that the thrust of advanced technology development in the past decade has involved a new kind of collaboration often described as the **Web 2.0** approach. It is:

- *technology neutral*, because it is not based on any particular software/hardware platform,
- *anonymous*, in that collaborators do not necessarily know one another,
- *bottom-up*, in that individual users add value to the system, and
- *responsive* – able to reflect change instantly and universally.

Consider a single, simple example: going *beyond* the dictionary definition of a word to explicate and exemplify its subtler uses and meanings. It is easy to imagine a Web page that presents a basic definition called from one or more Title VI-funded works, then lets the student click on an array of ancillary materials, including those produced by **other federally funded projects**, as well as other materials contributed by individual users:

- an *image* of the word as spray-painted on a wall or sign,
- *sound recordings* of the word in citation form, or in an authenic spoken context,
- a snippet of a YouTube *video* in which somebody uses the word,
- a listing of its *contexts and collocates* as found in a text corpus or on the Web itself.

We note in closing that the project does not duplicate past efforts to address these questions. Within Title VI, these have taken a traditional, library-oriented approach to the problem: identifying and evaluating results, and /or cataloging project outcomes, and include:

- *International Resource Information System* (IRIS), the Title VI project tracking system,
- *Education Resources Information Center* (ERIC), a centralized US/ED database of both metadata and some project outcomes (usually PDF files),
- *Language Materials Project* (LMP), on-line bibliographic database of teaching and learning materials for less commonly taught languages,
- *e-LCTL Initiative: a National Project for US/ED Title VI Programs*, a database that includes digital-format language modules, learning objects, and distance-learning plans.
- *Evaluation of the Impact of the IRSP* (JBL Associates and Center for Applied Linguistics), a study of all IRS projects (that, with US/ED permission, is sharing data with us).

4. FORMULATION OF THE PROBLEMS AND KNOWLEDGE OF RELATED RESEARCH (10)

4.1 FORMULATION OF THE PROBLEM

Our overarching concern is the development of a technology infrastructure, built from interoperable resources, and delivered using Web-based services.

Problem statement: *What "best practices" will enable interoperable digital services that can deliver Title VI resources, and meet the needs of the Title VI stakeholder community*?

Addressing this problem requires research into the four areas originally outlined in Section 1, and discussed in detail below: *resources* (**RQ 1**), *capacity* (**RQ 2**), *services* (**RQ 3**), and *protocols* (**RQ 4**).

Resources / capacity We cannot begin to approach the problem of building a technology infrastructure without knowing what raw materials we have to work with. Nor can we effectively advise projects on what services they should provide without evaluating their capacity for follow-through – their technical abilities, and/or the restrictions on data – that might limit participation. These provide research questions 1 and 2:

RQ 1. What lexical resources have been produced by IRS, TICFIA, and other Title VI projects? What is their availability for access and reuse?

RQ 2. What is the present capacity of active and past projects to follow "best practices" in digitizing data and/or providing services?

Services / protocols Then, given this understanding, we must consider the two distinct dimensions implicit in making Web-based services available. First, *what* services (dictionary lookup? phonetic approximation? reverse definition search?) should be provided? Second, *how* should the services be provided? Which of the many alternative protocols best meet our needs? These are restated as research questions 3 and 4:

RQ 3. What services should be recommended as "best practices" for enabling interoperability, and providing access to Title VI-funded resources?

RQ 4. What specific technical protocols should be adopted as "best practice" for providing Web-based services?

4.2 KNOWLEDGE OF RELATED RESEARCH

We look at related research for each RQ in turn.

RQ 1. What lexical resources have been produced by IRS and TICFIA-funded projects? What is their availability for access and reuse?

Section **3.1** listed many of the current database and archive efforts:

- *International Resource Information System* (IRIS),
- *Education Resources Information Center* (ERIC),
- *Language Materials Project* (LMP),
- *e-LCTL Initiative: a National Project for US/ED Title VI Programs*,
- *Evaluation of the Impact of the IRSP* (JBLA, CAL).

As noted, however, these are broad catalogs of project goals and outcomes. They may identify the type, general content, and location of a resource (e.g. "Chinese dictionary" at a library or project website), but have not been concerned with issues of interoperability and/or reuse.

RQ 2. What is the present capacity of active and past projects to follow "best practices" in digitizing data and/or providing services?

The dearth of information (or any way to find information) on this question is extreme. Yet it is critical that we do not operate in a vacuum, as the promulgation of standards which are unassailable in theory, but unacceptable in practice, is a longstanding IT industry issue.

We will address **RQ 2** by identifying, contacting, and interviewing the *technical and communications experts* who will ultimately be responsible for following "best practices" in digitizing data and/or providing services. In doing so, we consciously follow the advice given by the National Research Council:

> "However, a number of problems prevent the effective use of technology in language teaching for Title VI/FH programs … An integration of all available tools might lead to an effective language learning system, but such integration cannot happen easily, because it requires smooth data communications and interface synchronization. This is not possible if technical and communications experts are not a part of the team creating the language learning and teaching platform." (NRC, pp 241-242)

RQ 3. What services should be recommended as "best practices" for interoperability?

Many extant examples will inform our work, including:

- ad hoc services: practices followed by *Digital Dictionaries of South Asia, Tibetan-Himalayan Digital Library, Southeast Asian Languages Library, Nahuatl Learning Environment, SEASite*, and other projects.
- formal protocols: typical is the *DICT Protocol* (RFC 2229), published in 1997. It provides one basic query, **define**, and the possibility of various **match** strategies: *exact* and *prefix* are required, and a variety of others are allowed optionally.
- resource APIs: typical is the *WordNet* dictionary, which supports a rich set of queries that exploit its internal structure. WordNet projects world-wide follow essentially the same structure, and can be explored using the same kinds of queries.

These examples point the way, but they do *not* address our needs for a set of services that is well-defined (unlike the ad hoc services), rich in features (unlike Dict-like protocols), and is not tied to any one underlying implementation (unlike WordNet).

RQ 4. What protocols should be adopted as "best practices" for Web-based services?

The services described above will list the sorts of information that should be made available. However, there is a practical side to providing Web services as well; it describes the lower-level mechanisms by which users (or programs acting on their behalf) identify themselves to service providers (e.g. servers), as well as the means that providers use to respond.

In recent years, there has been a steady progression away from systems that require some sort of negotiation between the user and the service provider (such as the Z39.50 protocol, still in wide use for providing library services), and toward "lightweight" protocols that are intuitive, do not require storing the state of the current interchange, and can provide results in an immediately (re)usable form. We will consider both traditional (SOAP, WSDL) and modern, so-called **restful** approaches. Note that these are not mutually exclusive, and that it is possible to serve more than one protocol, and/or to 'translate' between protocol forms – indeed, preparing for and accommodating this is the thrust of "best practices."

5. Specificity of Statement of Procedures

We will follow the following procedures in addressing the four research questions.

Current data resources and capacity

- survey and summarize Title VI outputs (as described under **RQ 1, RQ 2**),
- inventory extant materials by type, condition, accessibility, needs, reuse restrictions,
- identify and contact "data custodians" (project managers),
- survey and/or interview "data managers" (project technical leads) for technical capacity,
- publish results for comment, correction, and dissemination.

Best practices for services

- survey and summarize typical services, dictionary-oriented protocols, and resource-led capabilities (as described under **RQ 3**),
- consult with project collaborators in refining "best practice" options,
- consult with data custodians and managers,
- propose and publish "best practices" for reference implementation, below.

Best practices for protocols

- survey and summarize typical protocols (as described under **RQ 4**),
- consult with project collaborators in refining "best practice" options,
- consult with data custodians and managers,
- propose and publish "best practices" for reference implementation, below.

Effectiveness and acceptability of "best practices"

- implement a reference set of basic dictionary query tools that rely on data providers' adherence to "best practices,"
- seek comment from collaborators, data custodians, and managers.

A typical "current data resources" survey instrument is shown below. It is a rough approximation of a Web-based survey form, in which alternative responses may be provided in drop-down lists or text boxes.

Name of project	(the formal, funded project name)
Name of resource	(the resulting dictionary or resource)
Funder details	(who paid for it)
Project leader	(Principal Investigator or Project Director and contact info)
Technical contact	(Past / current technical lead, if any)
Type of resource	(e.g. monolingual, bilingual, comparative, grammatical, etymological…)
Language	(list multiple languages if appropriate)
Current format	(paper, desktop publishing file, database, tagged file)
… type	(DTP file type / database name/version, tagging protocol/DTD)
… font encoding	(e.g. Unicode, one-byte national standard, self-defined)
… font availability	(are fonts for this encoding freely available?)
… sample entry	(if appropriate, paste a sample entry here)
Features	❑ definition ❑ romanization ❑ phonetic ❑ part of speech ❑ register ❑ subject ❑ etymology ❑ derivation ❑ morphology ❑ classifier ❑ auxiliary ❑ usage notes ❑ cultural notes ❑ synonyms ❑ antonyms ❑ equivalents ❑ intensifiers ❑ phrasal examples ❑ sentence examples
Shortcomings	(is the work complete? are parts missing?)
Data issues	(is the data corrupted or otherwise problematic?)
IP owner	(e.g. P.I., publisher)
Terms of release	(e.g. fair-use only, copy-left, open-source, Creative Commons, etc.)
Parent work	(is this data derived from a larger effort?)
Child work	(have derivative works based on this resource been produced?)
Equivalent work	(has the work appeared in other print or electronic form?)

Table 1 This table roughly approximates a Web-based survey instrument. Alternative responses may be provided in drop-down lists or text boxes to minimize effort. The finalized form can be submitted automatically.

6. ADEQUACY OF RESEARCH METHODOLOGY AND SCOPE OF PROJECT

6.1 ADEQUACY AND PROPRIETY OF RESEARCH AND TESTING

Procedures outlined in Section **5** have four focal points. Appropriate methods are used for each:

- *Current data resources and capacity* This will be done by survey, using Web-based forms, with initial contact and follow-up by both e-mail and telephone. We have done a considerable amount of advance work (see, for example, our preliminary survey of nearly one hundred primarily IRS-funded projects at http://sealang.net/temp/dictlist.htm), and foresee no difficulties.
- *Best practices for services* We will research and summarize existing practice, followed by discussion and consultation among project participants. As dictionary developers and/or service providers themselves, all project collaborators are familiar with the issues involved, and agree this is the best approach to establishing "best practices."
- *Best practices for delivery protocols* Research and summarize existing practice. Again, nearly all of the collaborators on this project are already dealing with this issue in one way or another, and agree that research, presentation, and discussion will lead us to agreement on "best practices."
- *Effectiveness and acceptability of "best practices"* A simple demonstration involving the DDSA, THDL, and SEAlang projects has already been established as a starting point (*Click! TICFIA!*). This will be expanded, used as a demonstration and focal point for discussion, and ultimately serve as a reference set that elucidates the "best practices."

By design, "best practices" are open-ended, and in all cases, we anticipate and welcome the opportunity for involvement and commentary by the Title VI developer community.

6.2 SIZE, SCOPE, AND DURATION OF THE PROJECT

Establishing an appropriate scale that would allow us to provide meaningful research results was a key factor in project design. We considered both broader and narrower formulations. For example, the larger question of preservation and access of digital resources in general in Title VI

must be considered at some point, but they involve too many technical issues in too many contexts – text, audio, image, video – for us to reasonably handle.

In contrast, once we had decided to focus on dictionaries, limiting the project to the IRS program would have imposed an artificial and unnecessary constraint, since TICFIA and the LRCs also work actively in this area, and deal with precisely the same issues.

Thus, in seeking collaborators, we have assembled a broadly representative group, with participants from a variety of funding programs (IRS, TICFIA, FIPSE, LRC) who have produced (and use) a very wide range of dictionaries, using an equally wide range of technologies. from print media in the 1970's to the present-day born-digital resources. Our **Advisory Board** is largely drawn from these (see Table **4** in section **8.2**, below); typical letters of support and commitment are also attached in the Apppendix.

The projects themselves range from what might be described as sole proprietorships established by individual linguists, to large-scale, well-entrenched, University-based library projects, for which dictionaries are just one of many services.

We have also **sought out diverse approaches** to current technology use. Our three largest collaborators – the DDSA, THDL, and SEAlang projects – make dozens of dictionaries available on-line, but have used different methodologies for everything from data tagging to Web hosting.

The project's three-year duration is essential. As the plan of operation in Section **7** details, we split the project into four nine-month phases: *acquisition* of data on projects and practices, *implementation* of our recommendations and demonstration sites, *dissemination to* – and more importantly, *feedback from* – participating projects, and finally *revision* and *outreach* to the broader stakeholder community.

It is crucial to recognize that stakeholder involvement and education is an integral part of the project. The data and recommendations that this project will deliver are, after all, only a means to an end: **increased access to taxpayer-funded resources**. Projects must have a sense of ownership and involvement in our work if they are to buy into the "best practices" we produce. And this takes time.

7. PLAN OF OPERATION

7.1 DESIGN

The 36-month project requires four phases: *acquisition, implementation, dissemination, revision*.

"Acquisition"	Survey existing project resources and development capacity (**RQ 1, 2**)
Months 1-9	Study existing "best practices" for interoperable services. (**RQ 3, 4**)
"Implementation"	Draft initial "best practices" recommendations. (**RQ 3**)
Months 10-18	Prototype initial tools / platforms for discussion.
"Dissemination"	Disseminate results to project members and other stakeholders.
Months 19-27	Survey and review acceptance / implementation issues. (**RQ 4**)
"Revision"	Draft final "best practices" recommendations. (**RQ 4**)
Months 28-36	Open access to demonstration services, release reference set of demonstration tools / platforms.

Action \ *Time, by quarter*	**Year 1**				**Year 2**				**Year 3**			
Acquisition	**Phase 1**											
Implementation				**Phase 2**								
Dissemination							**Phase 3**					
Revision										**Phs. 4**		
Evaluation												
Reporting												

Table 2 Project timeline. For an expanded timeline, see Appendix Table A1.

7.2 MANAGEMENT AND ADMINISTRATION

The *Center for Research in Computational Linguistics* and the Principal Investigator have considerable experience in managing resource development projects, and in the past five years have initiated 5 federally funded projects (total budget $1.8 million). The management plan is described here very briefly, and discussed further in Section **7.4**, below.

Administratively, the Principal Investigator has responsibility for grant management, for hiring and authorizing payment to project employees and consultants, for responding to requests made in the course of evaluating the project, and for fulfilling US/ED reporting requirements.

The PI, Project Manager, and IT Manager all work together closely, and will check progress on the project on at least a weekly basis. The primary collaborators will meet face to face at least once per year. Note that all outside collaborators manage projects that have both US and overseas-based components. Collaboration at a distance is commonplace, and we are all comfortable using modern methods, from ultra-low cost Skype conference calls to collaborative on-line editing environments like Google Docs.

7.3 RELEVANCE TO THE PURPOSES OF THE IRS (SECTION 605) PROGRAM

Section **3**, above, has already noted relevance to Title VI part A, which refers to development of specialized materials (1a), the use of advanced technologies for promoting access and dissemination of materials (2), and coordination between programs (3).

The same ties bind us to the purposes of Section 605, including references to research on **more effective methods** (*b.1*), development of **specialized materials** (c), studies and surveys of **use of technology** (*i*), and studies of practices for **dissemination of materials** (*j*). Three new authorized activates have been added in HEOA 2008; we directly address two (emphasis added):

> (*l*) *the systematic* ***collection, analysis, and dissemination of data*** *that contribute to achieving the purposes of this part; and*
>
> (*m*) *support for programs or activities to* ***make data*** *collected, analyzed, or disseminated under this section* ***publicly available*** *and easy to understand.*

7.4 USE OF RESOURCES (*see also Table 3, below*)

Principal Investigator: overall responsibility for RQ 1 – 4: set project quarterly goals, confer regularly with project collaborators, write "best practice" guidelines and reports, publicize work of the project at conferences and on line, assess the capacity of survey respondents,

Project Manager: identify and contact existing resource developers, draft survey instruments, manage the actual surveys (RQ 1 and 2). He will gather and help evaluate current practices of similar funding agencies (NSF, NEH, etc.), provide guidance to the project programmer, and disseminate project results to our stakeholders.

IT Manager: implement on-line survey instruments, provide technical consulting to survey respondents, test implementation of "best practices," implement demonstration tools.

Principal Investigator (0.25 FTE)	• establish production and review schedules, ensure timely project reviews, prepare semiannual reports for the IRIS project tracking system, • liaison between CRCL, DDSA, THDL, and other projects, • serve as lead software engineer in technical consultations / considerations.
Project Manager (0.5 FTE)	• monitor project progress in consultation with P.I. • draft instruments, manage surveys; evaluate "best practices" • prepare documentation, screencasts to assist in promulgating "best practices,"
IT Manager (project programmer) (0.5 FTE)	• implement on-line survey instruments; maintain office IT systems. • assist in analysis of respondents' data and capacity. • implement and test "best practices"; liaison with DDSA/THL IT staff
Office Manager (0.5 FTE)	• manage office/staff needs, assist Project Manager with surveys/dissemination.
Office Assistant (0.5 FTE)	• identify, contact, and follow up for survey and dissemination.
Reviewer (0.05 FTE)	• outside review of project progress, meeting of goals.
Collaborators (0.1 FTE)	• resource sharing, consultation on service capacity and requirements.
Digital Libraries / Software Consultant (40+40 hrs/yr)	• advise on (and design / install as needed) special-purpose software systems, • advise on website security and reliability issues,
Internationalization Consultant (40 hrs/yr 2-3)	• advise on internationalization issues; particularly those concerned with interoperability issues for non-Roman scripts.

Table 3 Responsibilities of project staff and primary consultants.

Office Manager / Assistant: manage survey, comment, dissemination contact and response.

Digital Library / Software Consultant: advise on existing and recommended "best practices" (RQ 3, 4); assist if necessary with reference implmentation.

Internationization Consultant: assist in meeting the Competitive Priority (see section **12**); in particular, interopability issus for non-Roman script LCTLs.

Collaborators: consult with the Principal Investigator on "best practices," assist in establishing reference implementation for Web services, provide data resources as necessary.

Reviewer: implement plan for periodic and summative reviews (Section **10**).

7.5 Equal access

The project is designed to help establish standards that will make taxpayer-funded resources as accessible as possible. It will expose data and services that were previously potentially **difficult to access**, and allow them to be **re-implemented and/or re-offered** in whatever manner is most suitable for providing equal access for all groups. This may include the development of user interfaces that require simpler Web access equipment, lower bandwidth, and/or present information in ways that accommodate those with limited visual or physical capacity.

8. QUALITY OF PERSONNEL

The Center for Research in Computational Linguistics (CRCL Inc, a US 501(c)3 nonprofit), is deeply involved in digital resource development. Our TICFIA-funded **Southeast Asian Languages Library** (2005-2009) project is providing digital bilingual dictionaries, and monolingual and bitext corpora for languages across Southeast Asia. Our IRS-funded **Southeast Asian Languages Laboratory** (2006-2009) project uses and extends these resources in researching and developing tools for learning complex-script languages. An NSEP-funded **Russian Digital Resources Project** (*Collaborative Innovation*, 2009-2012) applies these methods to the programs in The Language Flagship. And our **Mon-Khmer Languages Project** (2007-9, 2009-11) funded by the National Endowment for the Humanities, is digitizing data from the nearly 150 languages of this widespread family, building comparative and etymological data sets. CRCL works closely with the University of Wisconsin-Madison, and was chosen to host the 2009 TICFIA Project Director's Conference.

None of the CRCL projects work in isolation; every one is enabled by collaboration with research initiatives, academic institutions, and scholars around the world. This experience, and the profound lesson it has taught us regarding the fundamental importance of accessible, interoperable data resources, are behind this proposal.

8.1 PROJECT DIRECTOR

P.I. Doug Cooper is founder and Executive Director of CRCL. The author of nine college-level computer science texts, including the well-known *Oh! Pascal!*, he was previously a member of the Computer Science faculties of UC Berkeley and Smith College. He is a strong advocate of interoperability, and was instrumental in persuading Google to extend the SiteMap protocol to allow indexing of dictionary entries. His paper "**Digital Resources in Title VI: planning the next 50 years**" was presented at the Title VI 50th Anniversary Conference this year.

8.2 OTHER PERSONNEL

Project Manager Rikker Dockum is a graduate of Dartmouth College and a Fulbright Grant recipient. He originally joined CRCL in the pursuit of Thai-language research interests (in

particularly, lexicography and etymology), but has become our resident expert on regional language-resource issues, working on such tools as the SEAlang *WebCorpus*.

Digital Systems Manager Lwin Moe is a graduate of Purdue University, and received his MS in Computer Science from the Asian Institute of Technology. A native speaker of Burmese and Mon, Lwin manages CRCL's hardware and software systems. He has done a considerable amount of work in the area of Unicode standardization for Southeast Asian languages.

Digital Library Consultant Eric Lease Morgan is the Head of the *Digital Access and Information Architecture Department* of the University Libraries of Notre Dame. Eric consults for CRCL on a regular basis; we value his help in deciphering often obscure standards, and his breadth of understanding of long-term trends in digital library management and policies.

Internationalization Consultant Richard Ishida is *Internationalization Activity Lead* for W3C (the governing body of the *World Wide Web Consortium*), and co-chair of the Unicode Consortium's annual *Internationalization & Unicode Conference*.

Primary Collaborators James Nye and David Germano are singled out for comment because of their extensive experience, and because of the scope of their digital resource development and holdings. As Director of the Digital South Asia Library and its *Digital Dictionaries of South Asia* project Jim has shepherded some three dozen dictionaries through digitization, tagging, and on-line service. David, in turn, founded the *Tibetan and Himalayan Digital Library*, also the home of an extraordinary range of digital resources.

Evaluator Prof. Robert Bickner (University of Wisconsin-Madison) is a member of the advisory board of the University of Hawai'i *National Foreign Language Resource Center*, a founder and member of the *Consortium for Advanced Study of Thai* (which is establishing ACTFL-equivalent standards for Thai speaking proficiency), and is the P.I. of a US/ED-funded project to develop graded Thai-language reading materials.

Advisory Board An advisory board assures broad awareness of and participation in the project. Advisory Board members contribute both resources and expertise, and include the directors of US/ED and other federally funded projects (see Table 4, below).

James Nye	University of Chicago, *Digital Dictionaries of South Asia* (IRS, TICFIA)
David Germano	Virginia Tech, *Tibetan and Himalayan Library* (IRS, TICFIA)
Michael Maxwell	Director, Technology Use, *Center for Advanced Study of Language* (CASL)
Jonathan Amith	Gettysburg College, *Nahautl Learning Environment* (IRS)
Mohamed Maamouri	Linguistic Data Consortium, *Linguistically Complex Languages* (IRS)
Catherine Ingold	National Foreign Language Center, *LangNet* (FIPSE)
John U. Wolff	Cornell University *Dictionary of Cebuano Visayan* (US/ED)
Stephen O'Harrow	University of Hawai'i, *Interactive Intermediate Vietnamese* (IRS)
Helen Aristar	University of Eastern Michigan, *Linguistic Extension / Gold Ontology* (NSF)
Edna Andrews	Duke University, *Slavic & East European Language Resource Center* (LRC)
George Henry	Northern Illinois University, *SEAsite* (IRS, NSEP)
Eric Lease Morgan	Notre Dame, *Digital Access and Information Architecture Department*
Richard Ishida	World Wide Web Consortium (W3C), *Internationalization Activity*

Table 4 Advisory board for the project; some individuals head several projects.

8.3 TIME COMMITMENTS

Time commitments of project members are shown below.

Project Director	0.25 FTE
Project Manager	0.5 FTE
Digital Systems	0.5 FTE
Assistant	0.5 FTE
Digital Library	0.05 FTE
Collaborators	0.1 FTE total
Consulting (specialized software / internationalization issues)	0.1 FTE total
Evaluation	0.05

Table 5 Time commitments of persons associated with the project..

Based on our experience, this combination of inside staff and outside consultants and collaborators is appropriate to the project.

8.4 BROAD PARTICIPATION

CRCL is an equal opportunity employer, with a strong commitment to providing opportunities for traditionally underrepresented groups including women and members of ethnic minorities. Our core staff includes both men and women of Burmese, Thai, Lao, and Cambodian nationality.

9. BUDGET AND COST EFFECTIVENESS

9.1 ADEQUACY OF BUDGET

As noted above, CRCL has in the past five years managed five major federally funded projects, all dedicated to developing digital lexical resources. The budget we submit here is based on our experience in doing similar work under similar circumstances. Together with our collaborators and consultants, we are familiar with each element of the project, including the preparation and conduct of on-line surveys, the assessment and evaluation of technology and standards, and the development of model generic tools and common platforms.

We are well aware that projects which ultimately require input and some degree of cooperation from large numbers of participants never run quite as smoothly as planned. Thus, we anticipate and allow for the likelihood that we may have to provide extra measures of assistance in evaluating the current state of taxpayer-funded resources, and in assisting our collaborators in adapting their resources to the service architecture we ultimately propose.

9.2 REASONABLENESS OF COSTS

This project will pave the way for a technology infrastructure that dramatically enhances the accessibility, longevity, and usefulness of resources developed at a cost of many millions of dollars. The project's staff requirements – particularly the inclusion of technical personnel – and its three-year time frame reflect the investment required to attain this goal.

Although of necessity described in terms of surveys and research, the project is ultimately about **building a consensus** among practitioners in the field regarding the accessibility and reusability of data. It is essential that any recommendation be accepted as the result of a serious investigation of current practice, and of the particular needs and constraints of Title VI fundees.

This requires time and involvement. Past and present projects are widely dispersed; their directors are busy people, and their technical leads are not always readily available. Nevertheless, we must make every effort to involve them in our efforts in a meaningful way – not simply filling out questionnaires, but actively participating, and forming the critical mass needed to make "best practices" take hold as de facto standards.

10. EVALUATION PLAN

The project has developed a *program logic model* to guide evaluation. This approach is widely used for planning and evaluating publicly funded projects whose success cannot be measured in simple terms of profit or loss, and whose impact may not begin to be appreciated until after the project's term. The logic model addresses the requirements of the *Government Performance and Results Act* (GPRA), which requires all federal program to report measures of program performance, and is specifically recommended in the National Research Council's Title VI program survey (previously cited).

The review process is led by our evaluator, University of Wisconsin-Madison Prof. Robert J. Bickner, and joined by project collaborators. In reviewing the project, he will rely on the four overall guides – utility, feasibility, propriety, and accuracy – recommended by the American Evaluation Association.[3] The project will be evaluated as appropriate in conjunction with each of its four main phases – *acquisition, implementation, dissemination*, and *revision*.

Nominally these **interim reviews will be completed by project quarters** 3, 6, 9, and 12 (as discussed in Section 7.1); some parts of these will begin earlier in each phase. Results of this ongoing evaluation will form an integral part of our semi-annual IRIS reports.

Quantifiable metrics will be used whenever possible. However, it is important to note that not all project outcomes can be measured in strictly numerical terms. Much of the value of the program logic model accrues from its focus on *process* evaluation, feedback, and improvement, and the ongoing contribution it makes to project methodology. As implied above, there is often no way to adequately anticipate a project's long-term impact; but we can and must do all we can to assess and improve the manner in which it is conducted.

Below, we describe some of the **milestones** and **criteria** that Prof. Bickner will rely on in evaluating the project for the benefit of *a*) the project's PI and personnel, *b*) US/ED and the IRS program, and *c*) the larger Title VI community.

---

[3] From *The Program Evaluation Standards* (Joint Committee on Standards for Educational Evaluation, (1994). See http://www.eval.org/EvaluationDocuments/progeval.html .

10.1 IDENTIFICATION OF PARTICIPANTS AND STAKEHOLDERS

Participants and collaborators

- have participants and collaborators been contacted? are their roles in the project clear?
- has a schedule for meetings and/or consultation been established?
- have any promised contributions of data, software, or technical assistance been realized?

Stakeholder classification and constituency

- has the project clearly and adequately identified its stakeholder groups?
- have the special interests of different constituencies been elaborated?
- have methods for contacting/disseminating information to stakeholders been established?

10.2 SURVEY OF EXISTING RESOURCES AND PROJECT CAPACITY

Design of the survey instruments

- is the scope of the survey sufficient to guide future project activities?
- are the queries precise enough to elicit comparable responses?
- when possible, do queries allow numerically quantifiable responses?
- is there a mechanism provided to allow additional feedback from survey recipients?
- were stakeholders allowed input on the survey's form and design? on the survey pool?

Conduct of the survey

- were the survey instruments pilot-tested and improved as necessary?
- was the survey conducted in an effective manner? was the survey pool exhaustive?
- did subjects have sufficient time to respond?
- was there effective follow-up to ensure the broadest possible response?
- did the survey results provide the information it was intended to provide?

Evaluation of the survey response

- was data quantified or tabulated in a manner able to support meaningful inferences?
- was there a follow-up audit process? are the responses trustworthy? are they replicable?
- was there a quantifiable analysis of the survey pool? of actual respondents?
- did project collaborators and advisors participate in evaluating responses?

Dissemination of the survey response

- was the stakeholder community adequately identified and contacted?
- were survey results disseminated in an effective and timely manner?
- were survey results presented in a manner that was comprehensible to the stakeholders?

Survey follow-up

- was there a mechanism provided to allow feedback from stakeholders?
- were any portions of the survey modified and repeated in response to feedback?
- what concrete actions resulted from the survey?

10.3 STUDY OF CURRENT PRACTICES AND NEEDS FOR INTEROPERABILITY

Design of the study

- what specific questions is the study intended to address?
- is the scope of the study sufficient to guide future project activities?
- has an advisory committee been formed? has its guidance been sought?
- what published sources will be used? what informal sources?

Conduct of the study

- are the researchers well-informed? are they using appropriate sources?
- are both formalized and informal practices being considered?
- have agencies (NSF, NEH, IMSL) or institutions that may have conducted similar studies been contacted and consulted?
- have project stakeholders been given the opportunity to comment on the study? to contribute to it?

Evaluation of the study

- have appropriate consultants been identified to help clarify technical issues?
- has the project's advisory board been given the opportunity to join in evaluation?
- are alternative approaches suggested by the results clearly distinguished?

Recommendations for "best practices"

- have recommendations that meet the goals of the project been produced?

- have concrete examples been identified?
- have similar recommendations by other agencies (NSF, NEH, IMSL) been considered?
- has there been quantifiable participation in, and evaluation of, these recommendations?

Dissemination of "best practices"

- has the stakeholder community been identified and informed? who does it consist of?
- has sufficient information been provided to allow ready adoption? is additional documentation or education required?
- has a mechanism been provided for feedback, and re-evaluation if necessary?
- insofar as possible within the lifetime of this project, is there any quantifiable survey of the adoption of these "best practices"? of their incorporation into recommendations or funding requirements? of their feasibility and acceptance?

10.4 REFERENCE MODEL FOR TOOLS AND PLATFORMS

Specification of the tools / platforms

- what tools / platforms were selected for the reference model?
- were project advisors and collaborators consulted in the selection?
- do the tools / platforms involve project collaborators? to what extent?
- do the tools / platforms make appropriate use of the "best practices" recommendations?

Implementation of the reference model

- were the prototype tools / platforms properly implemented?
- do they effectively demonstrate the application of the "best practices" recommendations?

Dissemination and evaluation of the reference model

- how were they distributed and/or made accessible to the stakeholder community?
- was a quantifiable survey done of their functionality? of their effectiveness in demonstrating the project's "best practices" recommendations?
- insofar as possible within the lifetime of this project, is there any quantifiable survey of the use of the reference model's prototype tools / platforms?

11. ADEQUACY OF RESOURCES

11.1 FACILITIES

As noted earlier, the project will be produced by the Center for Research in Computational Linguistics (CRCL Inc, a US 501(c)3 nonprofit). In the past five years, our proposals have been awarded more than $1.8 million in federal grants by US/ED, the NEH, and NSEP. We also do industry consulting and development consistent with our nonprofit purpose, these projects range from improving the quality of Thai optical character recognition and Lao text messaging, to writing Burmese, Khmer, Thai, and Lao avian flu disaster preparedness glossaries.

CRCL also hosts the resources of a variety of other projects. These include the websites of two important linguistics publications (the *Mon-Khmer Studies Journal* and the *Journal of the Southeast Asian Linguistics Society*), and many joint collaborations, including the *Proto Oceanic Lexicon* project (with Australian National University), the *Acquisition of Jakarta Indonesian* project (with the Max Planck Institute), the *Loan Words in Indonesian and Malay* project (with SOAS and the KITLV) and others.

11.2 EQUIPMENT AND SUPPLIES

CRCL offices support the projects mentioned above, and provide all typical services, including high-speed Internet, secretarial assistance, telephones, mail, and the like. Our computer facilities are excellent, with Linux, Windows, and MacOS platforms; these are supplemented by a dedicated leased server at our ISP (*GoDaddy.com*), which in turn is supplemented by one or more dedicated virtual servers that are instantiated as necessary. Physical servers are leased from GoDaddy.com, and virtual servers are provided by the Amazon *Elastic Computing Cloud* (EC2) and *Simple Storage Service* (S3) services.

12. COMPETITIVE PRIORITY

This proposal addresses Competitive Preference Priority 2:

> *Research, surveys, assessments, or studies focused on any of the following seventy-eight (78) languages selected from the U.S. Department of Education's list of Less Commonly Taught Languages (LCTLs)*

Specifically, we address the issue of interoperability of dictionary resources that:

- use non-Roman scripts and/or non-standard diacritics,
- are included in the list of 78 less commonly taught languages,
- will be provided by the project's collaborators, and
- will be included in / demonstrated by the reference set of applications we produce.

These languages – all named by US/ED, and among our collaborators' (SEAlang, DSAL, and THL) resources – include **Baluchi, Bengali, Burmese, Cambodian, Hindi, Kashmiri, Lao, Marathi, Nepali, Panjabi, Pashto, Persian, Tamil, Telugu, Thai, Urdu,** and **Vietnamese**.

Most – but not all – of these languages have well-defined Unicode scripts, but one-byte legacy encodings (often national standards) are still common. This has serious implications for interoperability, as the Internet was originally designed with one-byte Roman characters (e.g. the ASCII character set) in mind. Allowing multi-byte Unicode characters to appear in queries has presented serious problems for Internet applications, addressed by a series of sometimes inconsistent standards. For example, here are two alternative forms for the same Thai-language query; the second is **urlencoded** (and has been abbreviated to fit):

http://sealang.net/api?service=dict&query=บ้าน …

http://sealang.net/api?service=dict&query=%E0%B8%9A%E0%B9%89%E0%B8%B2 ...

Should "best practices" require one form? Both? Either? This appears to be an utterly simple question, yet organizations large and small – including, for example, Microsoft! – have had their on-line services undermined because they failed to consider the problem.

We are assisted by consultant Richard Ishida, who is the *Internationalization Activity Lead* for W3C (the governing body of the *World Wide Web Consortium*), and co-chair of the Unicode Consortium's annual *Internationalization & Unicode Conference*.

Project Narrative

Other Narrative

Attachment 1:
Title: Pages: Uploaded File: **1237-irs-appendix.pdf**

APPENDIX

A PROJECT TIMELINE A1

C LETTERS OF SUPPORT AND COMMITMENT A2

Dr James Nye *on behalf of Digial Dictionaries of South Asia (University of Chicago)* **A3**

Prof. David Germano *for the Tibetan and Himalayan Library (Unversity of Virginia)* **A4**

Dr. Catherine Ingold *on behalf of LangNet (National Foreign Language Center))* **A5**

Dr. Jonathan Amith *on behalf of the Nahuatl Learning Environment (Gettysburn College)* **A6**

Dr. Mohamed Maamouri *for the Georgetown University Press / Linguistic Data Consortium Digital Dictionaries for Arabic Project* **A8**

Prof. George Henry *on behalf of SEAsite (Northern Illinois University)* **A9**

Prof. Helen Aristar *on behalf of the LEGO and Gold Cyberinfrastructure projects (Eastern Michigan University)* **A10**

E CURRICULA VITAE A11

Doug Cooper A12
Prof. Robert Bickner A15
Eric Lease Morgan A18
Richard Ishida A21

TABLE A1 TIMELINE **A1**

APPENDIX A: PROJECT TIMELINE

This table extends the brief view of section **7.1**.

Timeline	**Year 1**				**Year 2**				**Year 3**			
Action												
Acquisition	**Phase 1**											
Implementation				**Phase 2**								
Dissemination							**Phase 3**					
Revision										**Phs. 4**		
Evaluation												
Reporting												
Principal Investigator	**25%**											
Project Manager	**50%**				**(in effect, one tech-savvy project manager)**							
IT Manager	**50%**											
Office Manager	**50%**				**(in effect, one highly competent staffer)**							
Assiatant	**50%**											
Evaluation												
Software Consulting	**40 hours/year**											
Digital Libraries	**40 hours/year**											
Internationalization					**40 hours/year**							
Collaborating Projects					**DDSA / THL**							

Table A1 Project timeline, showing milestones and distribution of effort.

APPENDIX B: LETTERS OF SUPPORT AND COLLABORATION

Dr James Nye *on behalf of Digial Dictionaries of South Asia (University of Chicago)* **A3**

Prof. David Germano *for the Tibetan and Himalayan Library (Unversity of Virginia)* **A4**

Dr. Catherine Ingold *on behalf of LangNet (National Foreign Language Center)* **A5**

Dr. Jonathan Amith *on behalf of the Nahuatl Learning Environment (Gettysburn College)* **A6**

Dr. Mohamed Maamouri *for the Georgetown University Press / Linguistic Data Consortium Digital Dictionaries for Arabic Project* **A8**

Prof. George Henry *on behalf of SEAsite (Northern Illinois University)* **A9**

Prof. Helen Aristar *on behalf of the LEGO and Gold Cyberinfrastructure projects (Eastern Michigan University)* **A10**

e-mail letters of confirmation have been received from all other Advisory Board members.

THE UNIVERSITY OF CHICAGO LIBRARY

1100 East 57th Street
Chicago, Illinois 60637

Bibliographer for Southern Asia

April 14, 2009

Doug Cooper
Executive Director, CRCL
820 Calle Pluma
San Clemente, CA 92673-2719

Dear Mr. Cooper:

I am writing to confirm participation of the Digital South Asia Library (DSAL) in your IRS Research project proposal, "Interoperable Digital Resources in Title VI." This extremely important project will, in the long run, make a major contribution to initiatives funded by Title VI programs.

DSAL has a large number of bilingual digital dictionary resources developed under our *Digital Dictionaries of South Asia* project, including works on Assamese, *Baluchi, *Bengali, *Hindi, *Kashmiri, Lushai, *Marathi, *Nepali, Pali, *Panjabi, *Pashto, *Persian, Rajasthani, Sanskrit, Sindhi, *Tamil, *Telugu, and *Urdu (I have starred the twelve languages from the USED LCTL list). Other resources in varying stages of production include Divehi, Dogri, *Gujarati, *Kannada, Khowar, Konkani, *Malayalam, Manipuri, Maithili, Newari, Oriya, Prakrit, *Sinhala, and Torwali (with four additional USED LCTLs). Most of these languages use non-Roman scripts, and your project's focus on working with these scripts is of special interest to us.

The Digital South Asia Library fully agrees to the points we discussed with you and with our colleagues in the Tibetan and Himalayan Library. We will:

- comment on and contribute to the design and content of research or survey materials, and on the methodology and conduct of research and surveys,
- participate in research into the types of services that interoperable resources should provide, and in consideration of the most appropriate protocols for supplying these services,
- assist in the developing and evaluating alternative "best practices" for interoperability in providing resource access, and help promulgate these "best practices" when complete,
- make the contents of our bilingual dictionaries accessible via an API (application programming interface), noting that the same copyright restrictions that hold when these data are obtained via the DSAL website will still apply, and
- participate in the development of a reference implementation that demonstrates the application of "best practices" in providing resource access via the API.

DSAL will receive $5,000 per year in project years 2 and 3 to help defray development costs, and may receive additional technical assistance from the project if necessary.

Best wishes as you advance this very important effort.

Sincerely,

James Nye
Director, Digital South Asia Library

TIBETAN STUDIES
HALSEY HALL P.O. BOX 400126

TEL 434-924-6728
FAX : 434-924-1467

April 7, 2009
Doug Cooper
Center for Research in Computational Linguistics
820 Calle Pluma
San Clemente, CA 92673

Dear Doug:

The Tibetan and Himalayan Library enthusiastically accepts your invitation to participate in the International Research and Studies project: "Interoperable Digital Resources in Title VI."

The THL (www.thlib.org) project has always been a strong advocate of open access and shared resources. The standards and best practices that you propose to research and develop are desperately needed and long overdue. We greatly appreciate your taking the lead on this issue, which has been much discussed but never acted upon.

At present, the THL site supports dictionaries for many Himalayan languages, including Dzongkha, Jerung, Nepali, Sanskrit, Thangmi, Tibetan, and Wambule. We are also actively involved in building new kinds of dictionary-based reference works, such as the Tibetan Place Dictionary and Tibetan Historical Events Dictionary. These are accessible through dictionary search pages, and are being integrated into our Tibetan Language Learning Portal, Literary archives, and other resources.

THL agrees to the basic points of collaboration that you circulated. We will:

- comment on and contribute to the design and content of research or survey materials, and on the methodology and conduct of research and surveys,
- participate in research into the types of services that interoperable resources should provide, and in consideration of the most appropriate protocols for supplying these services,
- assist in the developing and evaluating alternative "best practices" for interoperability in providing resource access, and help promulgate these "best practices" when complete,
- make the contents of our bilingual dictionaries accessible via an API (application programming interface), noting that the same copyright restrictions that hold when these data are obtained via the DSAL website will still apply,
- participate in the development of a reference implementation that demonstrates the application of "best practices" in providing resource access via the API.

In project years 2 and 3 the THL will receive $5,000/year to cover our development costs. We will consult with your team on technical issues if necessary. However, we want to be actively involved as soon as the project gets underway.

Best wishes,

David Germano
Director, Tibetan and Himalayan Library
Director, UVa Tibet Center
Director, SHANTI (Social Sciences, Humanities, and the Arts Network of Technological Initiatives)

NATIONAL FOREIGN LANGUAGE CENTER
UNIVERSITY OF MARYLAND

April 12, 2009

Doug Cooper
CRCL INC
820 Calle Pluma Ave.
San Clemente, CA 92673

Dear Doug:

Thank you for inviting the National Foreign Language Center and the LangNet project to be involved with your proposed International Research and Studies project for "*Interoperable Digital Resources in Title VI.*"

I have discussed the project with our technical staff, most of whom attended the presentations you made on your last visit, and who were extremely impressed by the reference tools you have already built using these same principles. We all agree that your proposed research, and the results it will produce, will be of great practical benefit to our work, and we are delighted to help out in any way that we can.

As your proposal notes, the LangNet project was itself orignally funded by the Department of Education *Fund for the Improvement of Postsecondary Education* (FIPSE). It provides a language learning support system, with interactive materials designed for student and professional learners who want to practice and maintain their target language reading and listening skills.

Although we offer written and audio materials for dozens of languages, there is no way at present that we can take advantage of the equally large number of dictionaries that have also been developed with Department of Education funding. The demonstrations that you have given in your talks here have made a very compelling case for the merits of the "interoperability" approach.

From our point of view, the relatively small investment your project will require is going to have a tremendous impact on foreign language instruction, especially for the less commonly taught languages. Let me know if there is any way that I can help further this essential project.

Sincerely,

Catherine Ingold
Director, National Foreign Language Center

5201 PAINT BRANCH PARKWAY • PATAPSCO BUILDING, SUITE 2132 • COLLEGE PARK, MD 20742
TEL 301.405.9828 • FAX 301.405.9829 • WWW.NFLC.ORG

LINGUISTIC DATA CONSORTIUM
3600 Market St., Suite 810, University of Pennsylvania, Philadelphia, PA 19104-2653 USA • Tel: (215) 898-0464 • Fax: (215) 573-2175 • email: ldc@ldc.upenn.edu

April 12, 2009

Doug Cooper
Center for Research in Computational Linguistics, Inc.
820 Calle Pluma Ave.
San Clemente, CA 92673

Dear Mr. Cooper:

Thank you very much for your invitation to participate in your International Research and Studies project:

No More Data Silos: Interoperable Digital Resources in Title VI

This proposal is extremely well-timed. I am currently the recipient of an instructional materials grant from the IRSP, *Development of Digital Dictionaries for Iraqi, Moroccan and Syrian Arabic*. This proposal is a collaboration between the Linguistic Data Consortium (LDC) of the University of Pennsylvania and Georgetown University Press (GUP) to create up-to-date lexical databases, with translations to and from English, for three dialects of colloquial Arabic. The databases will be used both for interactive computer access and for new print publications of dictionaries in Iraqi, Syrian and Moroccan dialects.

Our work, and the IRS program as a whole, are desparately in need of the standards and practices that your project will devise. I have also explored this problem at length while preparing the *Arabic* project proposal, and can testify that at present, there are no common standards that may be followed by IRS or Title VI projects.

The approach that you are taking (of providing dictionary services, rather than redistributing complete data sets) is also exceptionally promising. It does appear to be a very workable solution to the copyright restrictions that we have to deal with.

Do not hesitate to get in touch if I may be of assistance in helping to convince the Department of Education to fund this project, or in aiding the project once it is underway.

Sincerely,

Mohamed Maamouri
Director/ Digital GUP/LDC Dictionaries

Gettysburg
COLLEGE

Department of Sociology
& Anthropology

300 North Washington St.
Campus Box 412
Gettysburg, Pennsylvania
17325-1486

717.337.6189 f 717.337.6666
www.gettysburg.edu

12 April 2009

Doug Cooper
CRCL INC
820 Calle Pluma
San Clemente, CA 92673

Dear Mr. Cooper:

Thank you very much for your invitation to be an advisor on your proposed IRS project, *Interoperable Digital Resources in Title VI.* I gladly accept and look forward to sharing the Nahuatl language resources that I have developed should they be helpful.

As you know, my Nahuatl Learning Environment represents an effort to link a dictionary, reference/pedagogical grammar, and extensive corpus of digital recordings with time-coded transcriptions into a synthetic pedagogical and research resource while making available to students and scholars the primary field data that were used in the elaboration of each component.

As a recipient of three USED IRSP grants (two as PI and one as a senior researcher) and a participant in several Title VI initiatives, I am familiar with difficulties that such endeavors face. I have read your proposal and would like to add my enthusiastic support for the manner in which you address problems that I feel have adversely affected both the implementation of projects by past recipients and the sharing of results among intended target audiences. First, you propose to establish, through survey and research, a set of "best practices" for building Web-based lexical and image databases and text corpora. This will be invaluable to future recipients. They will more easily learn from past experiences culled and made more accessible through your efforts, summarized in a series of "best practices" that you will make publicly available. Second, you recognize and propose solutions for what you aptly call "data silos", isolated materials that are not easily discoverable or retrievable (i.e., lack proper dissemination) through present resources. This would be a very welcome development. In sum, your efforts will increase accessibility to past efforts and guide future efforts in the creation of online dictionaries, image databases, and corpora.

The “best practices” and increased discoverability that your project will establish will also be invaluable to me in my own work, which was funded by the National Science Foundation, the Ford Foundation, and the National Endowment for the Humanities, as

well as the USED programs already mentioned. This kind of cooperation and collaboration is long overdue, and I am happy to be a part of it.

Please get in touch when you have any further word on the project.

Sincerely,

Jonathan Amith

Director: Mexico-North Research Network Program on Endangered Languages and the Nahuatl Learning Environment
Research affiliate: Smithsonian Institution, Gettysburg College, University of Chicago

NORTHERN ILLINOIS
UNIVERSITY

DEPARTMENT OF COMPUTER SCIENCE
DEKALB, ILLINOIS 60115-2854
(815) 753-0378
FAX (815) 753-0342

April 14, 2009

Doug Cooper
Center for Research in Computational Linguistics
820 Calle Pluma Road
San Clemente, California 92673

Dear Mr. Cooper:

Thank you for the invitation to join the advisory board of your project:

Interoperable Digital Resources in Title VI

As we discussed by telephone, I would be happy to help if I can. Your project is quite interesting, potentially very useful, and techologically quite current, given recent – and growing – interest in Web services in general.

As you know, I am also working on an IRS grant – **SEAsite: Website Infrastructure Improvements for the Next Decade** – so this is a very timely proposal.

You asked me to briefly describe the resources that I have developed for SEAsite and before. Beginning in 1992, we have created a series of bilingual dictionaries and learning tools for seven Southeast Asian countries. Many of the languages we work on use non-Latin scripts, which we are now revising in Unicode. We are especially interested in working on standards for access to dictionaries that use these scripts, which have all been mentioned in the Department of Education's list of 78 Less Commonly Taught Languages.

Please contact me if I can provide additional assistance or information to this important and long-overdue project.

Sincerely,

George Henry

George Henry
Associate Professor
Computer Science Department
Northern Illinois University

EASTERN MICHIGAN UNIVERSITY

April 23, 2009

Doug Cooper
CRCL Inc.
820 Calle Pluma Road
San Clemente, CA 92673-2719

Dear Doug:

I'm writing to formally confirm that I will join the Advisory Board of your "Best Practices for Interoperable Resources" project.

As you know, I am the P.I. of two related projects sponsored by the National Science Foundation: *LEGO: Lexicon Enhancement via the GOLD Ontology* (BCS-0753321, 2008-2011), and *Implementing the GOLD Community of Practice: Laying the Foundations for a Linguistics Cyberinfrastructure* (BCS 0720122, 2008-2009). Both of these encourage data interoperability through the use of GOLD – the General Ontology for Linguistic Description.

I have very much enjoyed the discussions we have had regarding our alternative approaches to building interoperable dictionaries, as well as the extremely impressive proof-of-concept demonstrations you gave us in your day of invited presentations and consultation at the *Institute for Language Information and Technology* here in Michigan last year.

In my opinion, your project is well designed, and the approach it takes is spot on in both technical and practical terms. It is an innovative, and entirely complementary consideration of the same critical issues that we address, and both of our projects will be well-served by discussion and cross-fertilization. I look forward to working with you on this important project, and will be happy to assist in any way that I can.

Sincerely,

Helen Aristar-Dry
Professor of Linguistics
Director, Institute for Language Information and Technology

Institute for Language and Internet Technology • 2000 Huron River Drive • Ypsilanti, Michigan 48197
Phone: 734. 487 0144 Fax: 734. 482 0132

APPENDIX C: CURRICULA VITAE

Doug Cooper A12

Prof. Robert Bickner A15

Eric Lease Morgan A18

Richard Ishida A21

Biographical sketch for Doug Cooper (born 1953, U.S. citizen)

Professional Skills

Project conception, development, and management: in the past five years have organized, written, and administered grants awarded $1.8 million, involving both CRCL staff and collaborators world-wide; funders include NEH, US/ED, and NSEP.

Software design and implementation: rather skilled at conceiving, prototyping, and managing development of novel software systems; particularly interested in *assistive technology* applications.

Writing and editing; book design and production: wrote series of widely used (~1,000 universities worldwide) and influential (for style, focus on problem solving) computer science textbooks; most recently edited a series of dictionaries of pre-Angkor and Angkorian Khmer.

Linguistics and epigraphy: broad knowledge of Southeast Asian comparative and historical linguistics, extensive work in building on-line digital resources for both. Fluent/literate in Thai.

Teaching and presenting: taught very large undergraduate courses at top schools; now spend considerable time giving presentations on design / implementation of digital resources for language learning and research.

Positions

2003-	Founder and Executive Director of the *Center for Research in Computational Linguistics* (CRCL Inc, a US 501(c)3 nonprofit); Director or Co-Director for all funded projects.
2008-	Adjunct, *Center for Advanced Study of Language*, University of Maryland
2007-	Honorary Fellow, University of Wisconsin-Madison
1995-2003	Headed informal *Center for Research in Computational Linguistics, Bangkok.*
1979-1993	Author, W.W. Norton & Co. (nine textbooks)
1989-1990	Lecturer, Computer Sconce Department, Smith College
1984-1988	Lecturer, Department of Electrical Engineering & Computer Science, U.C. Berkeley

Publications

(i) selected projects and systems

2009 **Russian Digital Reference Project**: dictionary, corpus, audio data for Russian.

2008 **Proto Oceanic Lexicon Project:** (with Prof. Malcolm Ross, ANU) comparative language database / etymological dictionary, novel map-based interface (see http://sealang.net/oceanic).

2007 **Mon-Khmer Languages Project:** broad project to develop languages database, etymological dictionary, and resource-sharing tool (see http://sealang.net/monkhmer)

2006 **SEAlang Lab:** assistive technology for reading, writing, and vocabulary acquisition in complex-script languages, project ongoing, see below and http://lab.sealang.net)

2005 **SEAlang Library:** on-line dictionary, corpus, and bitext corpus search tools (project ongoing, see http://library.sealang.net)

2003 **Keyster**: system for high-quality double-key entry of difficult texts by poorly educated operators (http://crcl.th.net/keyster/keyster.htm)

2000 **Proto-Tai'o'Matic:** on-line search system for proto-Tai and proto-Southwestern Tai family reconstructions (http://crcl.th.net/proto)

(ii) recent presentations and publications

2009 Editor, *Dictionary of pre-Angkorian Khmer* and *Dictionary of Angkorian Khmer* (Philip Jenner, Pacific Linguistics 2009)

2009 *Digital Resources in Title VI: planning the next 50 years*. Title VI 50th Anniversary Conference, Washington DC, 2009

2009 *the SEAlang Projects: Southeast Asian Language and Linguistics Resources*, UCLA – UC Berkeley Joint International Conference on Languages of Southeast Asia, Los Angeles, 2009.

2008 Invited presentations to the *Center for Advanced Study of Language* (CASL), the *National Foreign Language Center* (NFLC), the *Institute for Language Information and Technology* (ILIT), all in regard to projects at CRCL.

2008 Workshops for SEAlang Lab collaborators: the Foreign Service Institute, the Defense Language Institute, and the Center for Southeast Asian Studies, UW-Madison.

2008 *From My Land to Thailand: US Department of Education Initiatives for Thai Language Research and Education* (International Conference on National Language Policy, Royal Institute, Bangkok).

2008 *Sound[s|ed] like …? Approximate Phonetic Search in the Mon-Khmer Languages Project*. Southeast Asian Linguistics Society XVIII, Kuala Lumpur, Malaysia.

2008 *How to Speak ASEAN: Meeting the Challenge of Southeast Asian Language Research and Reference* (International Conference on Language, Literary Works, and Culture in ASEAN, Chulalongkorn University, Bangkok, Thailand).

2008 Presentations on SEAlang Projects to Southeast Asian Studies organization's annual meetings; including *SEASSI, COTSEAL, CORMOSEA,* and *SEAC.*

2007 *Research Tools for Complex Script Languages* (National Council of Less Commonly Taught Languages annual conference, Madison, Wisconsin).

2007 *Data Sharing in the Mon-Khmer Languages Project*. Third International Conference on Austroasiatic Linguistics (ICAAL 3), Puna, India 2007

2007 *The SEAlang Archives: Preservation, Discovery, and Access for the 'Scattered Literature' of Southeast Asian Linguistics*. Southeast Asian Linguistics Society XVII, Maryland, USA, 2007

2007 Workshops for SEAlang Lab collaborators: the Foreign Service Institute, the Defense Language Institute, and the Center for Southeast Asian Studies, UW-Madison.

2006 Assistant editor, *Mon-Khmer Comparative Dictionary* (H.L. Shorto, Paul Sidwell editor. Pacific Linguistics 2006).

2006 *The Southeast Asian Languages Library: On-Line Dictionaries, Text Corpora, and Tools for Southeast Asian Languages*. Council of Teachers of Southeast Asian Languages (COTSEAL) annual conference.

2005 *New Tools for non-Roman L2s: extensive reading, intensive vocabulary acquisition, and basic research.* Invited plenary speaker, Interagency Language Roundtable, Washington DC.

2005 *SEALANG Lab: 3/3/3+ Surge Capacity for Southeast Asian Languages*. Interagency Language Roundtable Showcase, Marylands.

2005 Invited presentations to the *Center for Advanced Study of Language* (CASL), the NSA, the *Linguistic Data Consortium* (LDC) and others re advanced tools for complex-script languages and the overall CRCL program.

1997 *How to Read Less and Know More: Approximate Optical Character Recognition for Thai* (Association for Computing Machinery SIGIR '97, Philadelphia, USA)

1996 *How Do Thais Tell Letters Apart? A Study of the Secondary Characteristics of Thai Letters* (Pan-Asiatic Linguistics, 4th Int. Symposium on Language & Linguistics, Mahidol University)

(iii) textbooks

1993 *Teaching Introductory Programming, 2nd Edition* (504 pages)

1993 *Oh! Pascal!, 3rd Edition* (lvi + 647 pages)

1992 *Oh! Pascal! for Turbo Pascal 6.0, 3rd Edition* (lxi + 836 pages)

1990 *Oh My! Modula-2!* (xxxii + 754 pages)

1987 *Condensed Pascal, with more Math and Science* (xxviii + 612 pages)

1985 *Teaching Introductory Programming* (x + 173 pages)

1985 *Oh! Pascal!, 2nd Edition* (lv + 540 pages)

1984 *Standard Pascal User Reference Manual* (iv + 183 pages)

1982 *Oh! Pascal!,* with Michael Clancy (xxx + 486 pages)

Synergistic Activities

1) Russian Digital Reference Project (Russian DRP) Funded 2009-2012 by the National Security Education Program under The Language Flagship *Diffusion of Innovation* program (sponsor: the Overseas Flagship / American Councils). This project will develop interoperable on-line resources and tools for Russian-language lexicography, and text, bitext, and audio corpora.

2) Mon-Khmer Languages Project Funded 2007-2009 and 2009-2011 by the US National Endowment for the Humanities (PM-50012-07), this project is assembling a century of data on the roughly 150 Mon-Khmer languages: linking it to modern comparative analyses, and making it accessible for research, reference and education. Assisted by leading scholars in the U.S., England, Germany, Australia, Singapore and Thailand, the project is building databases of language and etymological data, and a collaborative on-line worksite that manages a collaborative architecture for access, analysis, comment, and correction of database and dictionary data.

3) SEAlang Lab Funded 2006-2009 by the US Department of Education *International Research and Studies* program (P017A060058); project develops assistive technology for reading, writing, and vocabulary acquisition in complex-script languages. These include the *Reader's Workbench* (word and phrase segmentation, automated phonetic transcription, highlighting of core-vocabulary coverage, evaluation of text difficulty), the *Writer's Workbench* (addresses mechanical ability using predictive completion from local script, phonetic transcription, or letter-by-letter transliteration, and syntactic / expressive competence via integrated corpus and collocational reference tools), and the *Vocabulary Workbench* (uses novel heuristics that draw on dictionary and corpus data).

4) SEAlang Library Funded 2005-2008 by the US Department of Education *Technological Innovation and Cooperation for Foreign Information Access* program (P337A050018, University of Wisconsin-Madison lead institution), co-sponsor CRCL Inc. The project provides language reference materials for Southeast Asia, with an initial focus on the non-roman script languages including Thai, Burmese, Khmer, Lao, Mon, Shan, and Karen, as well as Vietnamese. Materials include bilingual and monolingual dictionaries, monolingual text corpora and aligned bitext corpora, and a variety of tools for manipulating, searching, and displaying complex scripts.

5) Tai and Tibeto-Burman Languages of Assam Funded by Hans Rausing Endangered Languages Project (P.I. Dr. Stephen Morey, Research Centre for Linguistic Typology, Latrobe University), with assistance from CRCL Inc. Project entails documenting a variety of endangered languages. CRCL has created a variety of support tools for on-line text corpora and dictionary development.

6) Southeast Asian Linguistics Archives (SEAlang SALA) Funded by CRCL Inc., this project collects, scans, indexes, disseminates, and analyzes the relationships between scholarly publications on Southeast Asian language and linguistics. The SALA includes a searchable on-line archives, an extensive bibliography, and most importantly a corpus of citations and analytical citation data. See sealang.net/sala.

Other Affiliations

The **SEAlang Lab** has cooperative agreements with the U.S. Foreign Service Institute (FSI), the Defense Language Institute (DLI/FLC), and the UW-Madison Center for Southeast Asian Studies. The **SEAlang Library** has agreements for cooperation / collaboration with various Southeast Asian institutions and organizations.

Robert J. Bickner
Dept. of Languages and Cultures of Asia
University of Wisconsin
Madison, WI 53706
(608) 262-3915, 265-3538 (fax)

Employment

1994 - present	Professor (Thai Language and Literature) and Chair Department of Languages and Cultures of Asia University of Wisconsin – Madison Principal Investigator, SEAlang Library (TICFIA 2005 – 2009) Principal Investigator, Thai Reading Materials Project (IRSP 2006 – 2009)
1989 - 1994	Associate Professor (Thai Language and Literature) Department of Languages and Cultures of Asia University of Wisconsin - Madison
1983 - 1989	Assistant Professor (Thai Language and Literature) Department of Languages and Cultures of Asia University of Wisconsin - Madison
1981 - 1983	Visiting Assistant Professor (Thai Lang. and Lit.) Department of Languages and Cultures of Asia University of Wisconsin - Madison
1970 - 1974	U.S. Peace Corps Volunteer Instructor of EFL and Teaching Methodology Chombung Teacher College, Ratchburi, Thailand

Education

PhD	1981	University of Michigan, Major field: Linguistics
MA	1977	University of Michigan, Major field: Asian Studies
BA	1969	Catholic University of America, Major field: English Literature

Language Competence

Fluent in Thai, some competence in, or abstract knowledge of: Cebuano, French, Hmong, Khmer, Lao, Latin, Sanskrit, Vietnamese.

Committee & Service Work

1982-present	College Year in Thailand Program (CYIT), Director
2000-present	Southeast Asian Studies Summer Institute (SEASSI), Language Director
1993-2000, 2007-2010	Chair, Department of Languages and Cultures of Asia, UW-Madison

Books and Monographs

1991 *An Introduction to the Thai Poem "Lilit Phra Law" (The Story of King Law)*, Monograph Series on Southeast Asia, Center for Southeast Asian Studies, Northern Illinois University, 254 pages.

Electronic Media Materials

1996 *Computer Assisted Instruction for Thai Reading*, a 32 lesson Macintosh HyperCard Set

Articles

2002 "Reflections on a Literary Dispute Between Jit Phumisak and Phra Worawetphisit," *Proceedings of the 10th Annual Southeast Asian Linguistics Conference of the Southeast Asian Linguistic Society 2000. Arizona State University Program for SEA Studies, Monograph Series Press*, 87-96.

1992 *"Some Textual Evidence on the Tai Sounds *ai* and *a~~u~~," Papers on Tai Languages, Linguistics, and Literatures In Honor of William J. Gedney on his 77th Birthday*, Carol J. Compton and John F. Hartmann, eds., Occasional Paper No. 16, Center for Southeast Asian Studies Northern Illinois University, 223-230.

---- "The Nature of 'Standard' Thai," (with Thomas John Hudak), *Journal of South Asian Literature*, 25.1, 163-175.

---- "The Problems of Southeast Asian Language Instruction: The Case of Thai," in *Southeast Asian Studies in the Balance: Reflections from America*, Charles Hirschman, et al., eds. Association for Asian Studies, 93-123.

1989 "Directional Modification in Thai Fiction: The Use of 'Come' and 'Go' in Text Building," David Bradley ed., *Papers in South - East Asian Linguistics* No. 11: *Southeast Asian Syntax*, 15-79. *Pacific Linguistics*, A-77, The Australian National University.

---- "Cultural Variation in Reflective Writing," (with Patcharin Peyasantiwong) *Writing Across Languages and Cultures: Issues in Contrastive Rhetoric, Written Communication Annual Vol. II*, Alan C. Purves, ed., Sage Publications, 160-74.

1986 "Thai Tones and English Loanwords: A Proposed Explanation," *Papers from a Conference on Thai Studies in Honor of William J. Gedney*, Robert J. Bickner, et al., eds. Michigan Papers on South and Southeast Asia. Ann Arbor: Center for South and Southeast Asian Studies, University of Michigan, 19-39.

---- "Changing Perspectives on Language and the Poetic Arts in Thailand," *In Celebration of King Ram Khamhaeng of Sukhothai: Seven Hundred Years of the Development of the Thai Language*, J. Hartmann, ed. CROSSROADS, Center for Southeast Asian Studies, Northern Illinois University, 3.1, 104-117.

Books Edited

1988 *Selected Writings in Tai Linguistics by William J. Gedney*, R. Bickner et al, eds., Michigan Papers on South and Southeast Asian Studies. Ann Arbor: Center for South and Southeast Asian Studies, University of Michigan.

1986 *Papers from a Conference on Thai Studies in Honor of William J. Gedney*, R. Bickner et al, eds., Michigan Papers on South and Southeast Asian Studies. Ann Arbor: Center for South and Southeast Asian Studies, University of Michigan.

Presentations & Consulting

2004 “Minority Adaptation in Thailand,” Wisconsin Dept. of Instruction pre-departure seminar for 18 academic personnel traveling to Thailand to study refugee populations.

-- -- Evaluator for the 2004 *Tuttle Asian Language Publication Grant*, Tuttle Publishing.

2003 “Doing Business in Thailand,” Wisconsin Dept. of Commerce, International Division.

2002 Translation and Interpretation services for Lutheran Social Services and Dane County Circuit Court, case No. 02 TP 42.

2000 "Reflections on a Literary Dispute Between Jit Phumisak and Phra Worawetphisit," the 10Th Annual Southeast Asian Linguistics Conference of the Southeast Asian Linguistic Society, Madison, May 4th - 7th.

1997 “Culture and Religions of Southeast Asia as They Affect Contemporary Issues,” Wisconsin Rural Leadership Program Conference, LaCrosse WI.

1991 "The Nature and Observance of Holidays in the Buddhist Cultures of Southeast Asia," Invited presentation for Madison Metropolitan School District, teacher in - service training program, November 20th. The lecture was subsequently broadcast several times over local cable television.

---- "Opening Doors to Thailand," for a session entitled *Diversifying Destinations: Southeast Asia*, 44th International Conference on Educational Exchange, Denver, Nov. 6-8.

---- "Stereotypical Views: East and West and Cross - cultural Adaptation," UW - Madison Center for Southeast Asian Studies, Friday Forum, April 19.

---- "The Problems of Southeast Asian Language Instruction: The Case of Thai," panel on Southeast Asian language instruction, Association for Asian Studies, New Orleans, April.

---- "Issues and Goals for the Development of Instructional Programs in Southeast Asian Languages," one of four invited presenters for a round table discussion organized by the Consortium of Teachers of Southeast Asian Languages (COTSEAL), Association for Asian Studies, April.

1990 "The Problems of Southeast Asian Language Instruction: The Case of Thai," invited presentation for a conference entitled "Southeast Asian Studies and the Social and Human Sciences," cosponsored by the Association for Asian Studies and the Joint Social Science Research Council/American Council of Learned Societies Committee on Southeast Asia, with funding from the Johnson, Ford and Luce Foundations, at the Wingspread Conference Center, Racine, WI..

1989 "Language and Culture in Asia," conference on "Linguistics and Language Learning and Teaching" at Payap University, Chiang Mai Thailand.

---- "Student and Faculty Exchange Programs with the United States," Khon Kaen University, Khon Kaen, Thailand..

---- "Translation of the Thai Literary Classic: Lilit Phra Law", Mahasarakham Teacher College, Mahasarakham, Thailand, March 7th, 1989.

---- "Subtlety in Communication; English vs. Thai," for a conference on "Teaching English in Its Cultural Context," Prince of Songkhlaa University, Pattani, Thailand, March 22nd, 1989

1988 "Some Observations on a Thai Village Ritual," Slide/Lecture presentation, also given in 1981, '82, '83, '85, and '87 for various audiences and classes.

1986 "Plot and Structure: The Catalog as a Coherence Device in "Lilit Phra Law," Council on Thai Studies, Madison.

---- "The Traditional Poetry of Thailand," Asian Poetry Evenings Series, an invited presentation for the Asia Society, New York.

---- "Artistic Expressions of the Thai," Peoples and Cultures of Southeast Asia Seminar, Residential Weekend Seminar Series, UW - Extension, Madison.

1985 "Some Textual Evidence on the Age of the Thai Vowel Distinction **ai* vs. **aɯ*," 18th International Conference on Sino - Tibetan Languages and Linguistics, Bangkok.

1984 "Linguistic Analysis and Thai Literary Studies: the Need for a New Beginning," invited lecture, the International Symposium on Language and Linguistics, Chiang Mai University, Thailand.

1983 "Literature and Literacy in Thailand: a Discussion of Some Significant Misunderstandings," Council on Thai Studies, DeKalb.

---- "Literary Studies and Historical Reconstruction: The Case of Thai," 17th International Conference on Sino - Tibetan Languages and Linguistics, Seattle.

---- "The Literature of the Chakri Dynasty of Thailand," discussant for the Rattanakosin (Bangkok) Bicentennial Conference, DeKalb.

---- "Thai Literature and Historical Linguistics," invited presentation, Southeast Asian Program and Department of Anthropology, University of Chicago.

---- "Adaptation of Indic Verse Forms in Thai," Conference on South Asia, UW - Madison.

---- "Refugees and Adjustment: Background Information and Appropriate Strategies for Local Service Providers," Regional Conference for Returned Peace Corps Volunteers, Madison.

1981 "Indochinese Refugees in America: Linguistic and Cultural Problems," Teachers of English to Speakers of Other Languages (TESOL), Midwest Regional Meeting, Urbana.

1980 "The Position of *chan* Poetry in the Literary Heritage of Thai," Conference on Thai Studies in Honor of William J. Gedney, Ann Arbor.

Eric Lease Morgan

Head, Digital Access and Information Architecture Department
University Libraries of Notre Dame
Notre Dame, IN 46556
(574) 631-8604 • emorgan@nd.edu

Higher education

Master of Information Science, Drexel University (Philadelphia, PA), 1987

Bachelor of Philosophy, Bethany College (Bethany, WV), 1982

Positions held and academic ranks attained

Head, Digital Access and Information Architecture Department, University Libraries of Notre Dame, 2001-

CEO and President, Infomomtions, Inc., 2000-

Network Technologies Development Librarian, North Carolina State University Libraries, 1991-2000

Clinical Instructor, University of North Carolina - Chapel Hill, School of Library and Information Science, 1995 and 1996

Librarian, Catawba-Wateree AHEC, Lancaster, SC, 1989-1991

Distinctions, honors, awards

Recipient of the 2005 Foik Award ("Librarian Of The Year"), University Libraries of Notre Dame.

Recipient of the 2004 LITA/Hi Tech Award for Outstanding Communication.

Co-PI with Emory University, Virginia Tech, and Oregon State University on a National Science Foundation Digital Library grant called OCKHAM in 2003.

Recipient of the 2002 Bowker/Ulrich's Serials Librarianship Award.

"2002 Movers & shakers: people who are shaping the future of libraries," *Library Journal*, March, 2002.

Current Expert in the LITA/Top Technology Trends Committee, 2000-

Apple Library of Tomorrow (ALOT) grantee to write *Teaching a New Dog Old Tricks: A Macintosh-based World Wide Web Starter Kit Featuring MacHTTP and Other Tools*, 1995.

Grand Prize winner of the Meckler Computers in Libraries 1991 Computer Software Contest for QuickCat, an electronic card catalog, 1991.

National Library of Medicine (NLM) grantee to implement an information outreach program, 1990.

Apple Computer ALOT grantee to implement a medical librarian circuit-rider program, 1989.

Professional memberships and service

Co-Editor of *Code4Lib Journal* (2007-2008)

Official archivist for the ACQNET (2000-) and colldv-l (2005-) mailing lists. Moderator for the code4lib (2003-) , usability4lib (2005-), and ngc4lib (2006-) mailing lists.

Information Technology and Libraries (ITAL) Editorial Board Member, 2000-2002

United States Agriculture Information Network (USAIN) member from 1998-2001, mailing list manager from 1994-2001, and Communications Committee member from 1998-2001

North American Serials Interest Group (NASIG) member from 1998-2005

Library and Information Technology Association (LITA) member 1992-

American Libraries Association (ALA) member, 1989-

Selected articles and book chapters

Guest Editor of "Open Source Software in Libraries," *ASIS&T Bulletin*, In press.

"MyLibrary: A digital library framework and toolbox", *Information Technology and Libraries*, In press.

"Ethical and economic issues surrounding freely available images found on the Web", *First Monday* 11(7), July 2006, (http://firstmonday.org/issues/issue11_7/morgan/).

"Exploiting 'Light-weight' Protocols and Open Source Tools to Implement Digital Library Collections and Services", *D-Lib Magazine* 11(10), October 2005, (http://www.dlib.org/dlib/october05/morgan/10morgan.html)/

"Creating and Managing XML with Open Source Software", *Library Hi Tech* 23(4), 2005, pp. 526-540, (http://infomotions.com/musings/xml-with-oss/).

"An Introduction to the Search/Retrieve URL Service (SRU)", *Ariadne* (40), July 2004 (http://www.ariadne.ac.uk/issue40/morgan/).

"Open Source Software for Libraries," in Karen Coyle, ed., *Open Source Software for Libraries: An Open Source for Libraries*: Chicago: Amerian Library Association, 2002 pg. 7-18.

Guest Editor of "Special Issue: User-Customizable Library Portals," *Information Technology and Libraries* 19:4 (December 2000) (http://www.lita.org/ital/ital1904.html).

Libraries of the Future, a regular monthly column in *Computers in Libraries* (September 1997 - December 1999).

"Clarence Meets Alcuin or Expert Systems Are Still an Option in Reference Work" In Pat Ensor, ed., *The Cybrarian's Manual* Chicago: American Library Association, 1997 pg. 127-134.

"Possible Solutions for Incorporating Digital Information Mediums into Traditional Library Cataloging Services," *Cataloging & Classification Quarterly* 22:3/4 (1996) pg. 143-170.

"Description and Evaluation of the 'Mr. Serials' Process: Automatically Collecting, Organizing, Archiving, Indexing, and Disseminating Electronic Serials," *Serials Review* 21:4 (Winter 1995): 1-12.

Teaching a New Dog Old Tricks: A Macintosh-based World Wide Web Starter Kit Featuring MacHTTP and Other Tools (Apple Computer, Inc.: Cupertino, CA 1995)

"The World-Wide Web and Mosaic: An Overview for Librarians," *The Public-Access Computer Systems Review* 5(6):5-26, 1994. Japanese translation appears in *Research and Development of Scholarly Information Dissemination Systems Report for the Fiscal Year 1994* (Ministry of Education, Science and Culture: Japan) 66-84.

WAIS and Gopher Servers: Finding Your WAIS Through a Gopher Hole (Mecklermedia: 1994)

"A Day in the Life of Mr. D." in R. Bruce Miller and Milton T. Wolf, eds. *Thinking Robots, an Aware Internet and Cyberpunk Librarians* (Library and Information Technology Association: Chicago 1992) 151-156.

"Implementing TCP/IP Communications with HyperCard," *Information Technology and Libraries* 11(4):421-432, December 1992.

"The MARC Reader: a HyperCard Script for Demystifying the MARC Record," *Computers in Libraries* 11(11):52-55, December 1991.

Selected invited lectures and addresses

"Catalog Collectivism: XC and the Future of Library Search", Charleston Conference, Charleston, SC, November 8, 2007.

"Today's Digital Information Landscape: Plenty of Problems to Solve, Plenty of Solutions to Choose From", Texas Center for Digital Knowledge, University of North Texas, Denton, TX, December, 2007.

"Open Source Software for Libraries and XML", Digital Libraries à la Carte 2007, University of Tilburg, Tilburg, The Netherlands, August 29, 2007.

"Ethical and economic issues surrounding freely available images found on the Web", First Monday Conference, Chicago, IL, May 16, 2006.

"MyLibrary workshop", Annual Seminar of the Croatian University and Scientific Libraries, Zagreb, Croatia, March 2, 2006.

"'Light-weight' Protocols", OAI4 Conference, Geneva, Switzerland, October 21, 2005.

"Getting Started with XML: A Workshop" first facilitated for InfoPeople of California in 2003 and repeated numerous times across the country including the 2003 Museum Computer Network Annual Meeting in Las Vegas and the California State University – Monterey in 2004.

"'Light-weight Reference Models' in MyLibrary," European Conference on Digital Libraries, Rome, Italy, September 16, 2002

"MyLibrary@NCState," LITA National Forum, Raleigh, NC, November 7, 1999. This same presentation has been given at numerous other organizations including the Ontario Library Association, OCLC, the Canadian Library Association, LAUNC-CH, NERCOMP, New England Chapter of ASIS (Harvard), Fiesole Retreat (Oxford), WILSWorld (Madison, WI), SIGIR '99 and ALA Annual Meeting (Atlanta).

"Becoming a World Wide Web Server Expert," Library and Information Technology Association (LITA), Williamsburg, VA, March 14, 1997, Florida International University, Miami, FL, May 16, 1997, and University of Chicago, March 17, 1998.

"World Wide Web Publishing," Ties That Bind: Community Networking Conference, Apple Computer, Cupertino, CA, May 3, 1995.

Selected information systems and projects

University of Notre Dame Institutional Digital Repository (June 2006) – Uses OAI to harvest metadata from three distinct repository applications (Dspace, DigiTool, and ETD-db) to create a single interface to them all as well as provide enhanced services against the content. (http://www.library.nd.edu/idr/)

MyLibrary@Ockham (March 2005) – An OAI harvester and indexer demonstrating ways to statistically enhance metadata, improve user searches, and suggest ways to increasing retreival. (http://mylibrary.ockham.org/)

Ockham Alerting Service (January 2005) – A system for harvesting OAI or MARC data, saving it to a central store, indexing it, and providing access to the index through SRU. Search results can be disseminated via HTML, RSS, or email. (http://alert.ockham.org/)

My TEI Publisher (December 2004) – A system for creating and maintaning sets of TEI documents, transforming them into valid HTML files, and creating a searchable/browsable collection on the Web. (http://www.infomotions.com/musings/tei-publisher/)

Alex Catalogue of Electronic Texts (September 2000) - A collection of digital documents including items from American literature, English literature, and Western philosophy, allowing users to create customizable output in the form of PDF documents or ebooks as well as providing the means for extensive searchability of Catalogue content. (http://www.infomotions.com/alex/)

MyLibrary@NCState (February 1999) - A user-driven, customizable, portal-like interface to a library's collection of Internet resources providing patrons with current awareness services and direct access to librarians. (http://dewey.library.nd.edu/mylibrary)

Mr. Serials Process/Index Morganagus (December 1997) - A process for collecting, organizing, indexing, and disseminating library-related electronic serials. (http://sunsite.berkeley.edu/IndexMorganagus/)

The Alcuin Database (March 1995) - A database of more than 2,000 MARC records describing Internet resources, mounted using DRA software, and searched through a HTTP-DRA gateway script.

DocDoctor (September 1993) - A HyperCard stack developed for the NCSU Digitized Document Transmission Project allowing end-users to scan, compress, transmit via FTP, receive, uncompress, and print TIFF images.

HyperDialog (January 1992) - A HyperCard stack implementing a front-end to the DIALOG Information Service set of databases using TCP/IP as the communications medium.

Richard Ishida

BA Joint Hons., MPhil (Cantab)
ishida@w3.org, +44 1582 623 033, http:/ /rishida.net
Key passions: multilingual web, writing systems, education on a practical level
Key values: openness, respect for others, constructive team contributions, quality
Key skills: synthesis, simplification, presenting, listening, practicality, innovation

General Information

- **Unicode Standard** Editorial Committee member since 2008.
- Co-chair of the **Internationalization & Unicode Conference** from 2003-2005. Expanded conference remit to internationalisation as well as Unicode. On advisory board, 2005 to present. On review board since 2002. Presentations and tutorials at every Unicode conference since 1995.
- *Languages:* Native English; fluent French and Spanish; intermediate Japanese; also in-depth studies in German, Russian, Italian and basic Arabic.
- Age: 51. Married.

Employment

World Wide Web Consortium (W3C)
2005 – present Internationalisation Activity Lead
2002 – 2005 Technical Staff, Internationalisation

W3C is an organisation of some 55 staff based at host organisations in France (ERCIM), the US (MIT) and Japan (Keio University). Its mission is to develop base standards that support inter-operable information on the Web. Funded by more than 400 member organisations.

- Worked with and guided a wide range of technical *Working Groups* on internationalisation issues in emerging Web technologies.
- Regular contributions to internationalisation aspects of technologies such as HTML, XHTML, CSS, SVG, SSML, XML, WAI, etc.
- Added guidelines, education and outreach to the work of the Internationalisation Activity, and produced a large number of the resources that support that outreach.
- Presenter at high-visibility events worldwide, including @media conferences, tutorials at World Wide Web conferences, W3C 10th Anniversary event in France, numerous keynotes.
- Pioneered the concept of the *Internationalization Tag Set* (ITS), which resulted in publication of the *ITS Standard* and *Best Practices for XML Internationalization*.
- Chair and staff contact for several W3C *Working Groups* (which set international standards).

Xerox, Global Knowledge & Language Services (GKLS)
1996 – 2002 Globalisation Consultant
1994 – 1995 Senior Multilingual Design Specialist
1989 – 1994 Senior Translation Systems Specialist

Organisation of 800 people based in the UK, the U.S. and Singapore, providing services in document and web site creation, translation and localisation, printing and distribution.

- Numerous collaborations with industry bodies such as the Localization Institute and the Localization Industry Standards Association (LISA).
- Consultancy successes and proposals to senior management led the way in highlighting Globalisation as a key Go-to-Market strategy for the organisation.
- Key contributor to the team globalising the Xerox corporate product development process. Team earned major corporate award for contributions to Time-to-Market.
- Technical project management for product re-engineering to support Hebrew and Farsi translations, and feasibility study for Chinese and Korean - a new service for GKLS.

CADCentre Ltd.
1988 – 1989 Senior Software Engineer
Based in Cambridge, UK., develops industry-leading applications for designing large industrial plants.

- Software engineer, on a team developing a programmable command processor and compiler in FORTRAN and C.
- Rewrote company's Quality Assurance documentation (ISO 9000).

Other employment
1987 Research Assistant, University of Cambridge, analysis of speech waveform data.
1986 Temporary lecturer, University of Salford, Summer courses on translation and interpreting.

Education
1986 – 1987 MPhil, Computer Speech & Language Processing, University of Cambridge

- Speech recognition & synthesis. Natural language processing. Psychoacoustics & psycholinguistics. LISP and PASCAL programming.
- Thesis based on six month project developing a machine translation system.

1982 – 1986 BA, Joint Hons, Computer Speech & Language Processing, University of Salford

- Translating, interpreting. European studies. Linguistics & literature. Russian subsidiary.
- Persuaded the university to add new course modules in programming, phonetics and speech synthesis.
- 1 year residence abroad (France & Spain).

Service
Committees and Societies
2008–present. Unicode Standard Editorial Committee
2002–present. Review board, Internationalization and Unicode Conference.
2003–2005. Co-chair, Internationalization and Unicode Conference.

Editor for the following W3C publications:
Character Model for the World Wide Web: Fundamentals (W3C standard),
Internationalization Best Practices: Specifying the language of content 1.0,
Internationalization Best Practices: Handling Bidirectional Text 1.0,
Best Practices for XML Internationalization,
Japanese Layout Requirements.

Publication
Selected Technology Reports
2003. Requirements for Localizable DTD Design. (With Yves Savourel.)
2002. Localizable DTD Design. *MultiLingual* #49, Vol. 13, No. 5: Sandpoint, ID.
2001. Text Fragmentation and Reuse in User Interfaces. *MultiLingual* #43, Vol. 12, No. 7: Sandpoint, ID.
2000. Localisation Considerations in DTD Design. Xerox Global Services.

Selected Tutorials and Panels
2003. Moderator, Panel on Web Internationalization. 23rd Internationalization and Unicode Conference, Prague, Czech Republic.
2003. ”Designing International User Interfaces,” 2003 Multilingual Standard Globalization and Language Technology Architecture Seminars, San Francisco and Orange County, CA.
2000. "Designing International User Interfaces." Course, Localization Institute, Seattle, WA
1999. "Designing International User Interfaces." Course, Localization Institute, Stuttgart, Germany.

Selected Papers and Presentations

2007. "Hints for Designing International Web Pages." 31st Internationalization and Unicode Conference, San Jose, CA.

2007. "An Introduction to Writing Systems & Unicode." 31st Internationalization and Unicode Conference, San Jose, CA.

2006. "Where Would We Be Without UNICODE?" Internationalization and Unicode Conference, Washington, D.C. (with Mark Davis, et al.) Keynote address.

2006. "Internationalizing Style on the Web: What's Planned for CSS3." 30th Internationalization and Unicode Conference, Washington, D.C.

2006. "How to Express Information about Internationalization and Localization in XML Documents and Schemata." 29th Internationalization and Unicode Conference, Burlingame, CA.

2006. "An Introduction to Writing Systems & Unicode." 30th Internationalization and Unicode Conference, San Francisco, CA; also Internationalization and Unicode Conference 30, Washington, D.C.

2004. "Hot Web Schemas? SOAP to OWL." Localization World San Francisco 2004, San Francisco, CA.

2003. "Web Internationalization -- An Update from the W3C." (With Martin Durst.) 23rd Internationalization and Unicode Conference, Prague, Czech Republic.

2003. "Introduction to Indic Scripts." 24th Internationalization and Unicode Conference, Atlanta, GA.

2003. "Creating Bidi X/HTML Pages." 24th Internationalization and Unicode Conference, Atlanta, GA.

2003. "An Introduction to Writing Systems: A review of script characteristics affecting computer-based script support and Unicode." 24th Internationalization and Unicode Conference, Atlanta, GA.

2002. "Introduction to Indic Scripts." 22nd International Unicode Conference, San Jose, CA.

2000. "Non-Latin Writing Systems: Characteristics and Impact on Multinational Product Design." 17th International Unicode Conference, San Jose, CA.

2000. "Localization Considerations for DTD Design." 17th International Unicode Conference, San Jose, CA.

1999. "Non-Latin Writing Systems: Characteristics and Impact on Multinational Product Design." 14th International Unicode Conference, Boston, MA; also 15th International Unicode Conference, San Jose, CA.

1999. "Making Your Product Translatable." 14th International Unicode Conference, Boston, MA; also 15th International Unicode Conference, San Jose, CA.

Websites

Collected writings for W3C	http://www.w3.org/People/Ishida/writing
Personal site	http://rishida.net/
Blog on internationalisation issues	http://rishida.net/blog/
Other writings on internationalisation	http://rishida.net/writing

Software and Systems

Internationalisation Checker	http://rishida.net/rishida/tools/i18nchecker/
IANA Language Subtag Registry Search	http://rishida.net/utils/subtags/
Latitude and Longitude Converter	http://rishida.net/rishida/gps/convertlatlon.php
Territory Finder	http://rishida.net/rishida/tools/territories/
Byte Checker	http://rishida.net/tools/showbytes/
UniView Unicode database viewer	http://rishida.net/rishida/scripts/uniview/
Unicode Character Pickers (for 21 languages)	http://rishida.net/rishida/scripts/pickers/
Unicode Code Converter	http://rishida.net/rishida/scripts/uniview/conversion
UTF-8 BOM Tester	http://rishida.net/rishida/utils/bomtester/
Unicode Normalizer	http://rishida.net/rishida/tools/normalizer/

Budget Narrative

Budget Narrative

Attachment 1:
Title: Pages: Uploaded File: **1234-irs-budget.pdf**

Budget Narrative and Supplemental Budget Information

Overview

Budget details for each of the three years of the project are attached below.

Items that respond to the competitive preference priority

Our response to this priority involves addressing interoperability issues for non-Roman-script languages / languages listed in the US/ED LCTL list that are included among the resources of three primary collaborators: the Southeast Asian Languages Library (SEAlang), the Digital South Asia Library (DSAL), and the Tibetan and Himalayan Library (THL). Expenditures that directly correspond to this response include:

- most of the $5,000 / year / project in years 2 and 3 to DSAL and THL to defray costs associated with developing the reference applications that demonstrate out "best practices" (SEAlang effort is contributed).
- $3,000 / year in years 2 and 3 to the Internationalization Consultant (Richard Ishida).
- 10 – 20% of the $6,000 set aside yearly for software and digital library consulting.

Many of the dictionary projects and resources that we will survey also involve the languages named in the competitive priority, of course, and some portion of the overall cost of the proposal could reasonably be assigned to it.

Notes

1. **Salaries**: We request partial salaries (maximum 50%) for several staff positions. The P.I. position is 0.25 FTE, and the Project Manager, IT Manager, Office Manager, and an assistant are all 0.5 FTE. A yearly increase of 4% is factored into years two and three.

   Note that the P.I. is not compensated for his service as Executive Director of applicant CRCL Inc. He will have an existing 0.3 FTE obligation (to 0.55 if a pending TICFIA proposal is funded) at the start of this proposed IRS project.
2. **Fringe Benefits**: Rates are 34.5% (senior staff), 24% (junior staff); these rates are considerably lower than those of the academic institutions we partner with.
3. **Travel**: No travel expenses are included.
4. **Equipment:** No equipment expenses are included.
5. **Supplies:** No supply expenses are included.
6. **Contractual**: Fees for consulting, evaluation, and collaborating projects are included here. We estimate rates at $75 / hour.
7. **Construction:** No construction expenses are included.
8. **Other:** Costs for consulting, assisting collaborating projects, and project evaluation are included here, at an average total of $18,000 per year. Technical consulting rates ($75/hour) are extremely reasonable.
9. **Total Direct Costs**: $471,940.
10. **Indirect Costs**: We request an 8% indirect cost rate. This averages $1,000 / month. Note that we do not request remuneration for supplies or direct office expenses (e.g. rent, computing, telephones, etc.).

11. **Training Costs**: not applicable.
12. **Total Costs**: $509,695.

Other contributions

Although they do not appear on the balance sheet, the volutary assistance of the Advisory Board members, and the SEAlang Library, will make a significant in-kind contribution to the project. Similarly, travel costs incurred by the project (e.g. yearly meetings with DDSA/THL) are absorbed by other funded projects.

Budget, 2009-2012

U.S. Department of Education, IRS							
Interoperability for Digital Resources in Title VI							
Budget, October 1, 2009 - September 30, 2012							
		base	*pct*	**Year 1**	**Year 2**	**Year 3**	**Total**
1. Personnel	Principal Investigator (Cooper)	$128,000	25.0%	$32,000	$33,280	$34,611	$322,305
	Project Manager (Dockum)	$44,500	50.0%	$22,250	$23,140	$24,066	
	IT Manager (Moe)	$40,000	50.0%	$20,000	$20,800	$21,632	
	Office Manager	$36,000	50.0%	$18,000	$18,720	$19,469	
	Assistant	$22,000	50.0%	$11,000	$11,440	$11,898	
2. Fringe	Senior Staff		34.5%	$18,716	$19,465	$20,243	$95,135
	Junior Staff		24.0%	$11,760	$12,230	$12,720	
3. Travel				$0	$0	$0	$0
4. Equipment				$0	$0	$0	$0
5. Supplies				$0	$0	$0	$0
6. Contractual				$0	$0	$0	$0
7. Construction				$0	$0	$0	$0
8. Other	Digital Libraries	$75	40/40/40	$3,000	$3,000	$3,000	$54,500
(**priority-related*) *	Software Consulting	$75	40/40/40	$3,000	$3,000	$3,000	
*	Internationalization Consulting	$75	0/40/40	$0	$3,000	$3,000	
	Project Evaluation	$3,500	100%	$3,500	$3,500	$3,500	
*	Collaborating Projects	$5,000	0/2/2	$0	$10,000	$10,000	
10. Indirect Costs			8%	$11,458	$12,926	$13,371	$37,755
11. Training Stipends				$0			$0
TOTAL	*Total expenditures*			$154,684	$174,501	$180,509	$509,695