



## ***REQUEST FOR PROPOSALS***

### ***Joint Venture for the Cooperative Digitization and Dissemination of World Newspapers***

#### ***Phase One: Latin American Newspapers from the Era of Independence, State Formation, and Early Modernization, 1805-1923***

##### ***1. Rationale and Objectives of the World News Archive***

Through a century of major, sustained investment in acquisition, documentation and preservation, North American research libraries have amassed a large and valuable corpus of newspapers from all regions of the world. Those libraries' aggregate holdings of newspapers in paper and micro-formats constitute a body of historical and cultural evidence spanning four centuries, which is not, and could not be, replicated elsewhere.

The Center for Research Libraries and a number of major North American research libraries (here "Affiliates") now intend to undertake the systematic, large scale digitization of this corpus of world newspapers and news-related materials. The proposed initiative, here referred to as the *World News Archive*, is envisioned as an ongoing, multi-year and multi-stage endeavor wherein CRL and Affiliates will combine expertise and resources to digitize and make available for scholarly use, through the Web and other electronic means, their holdings of newspapers from seven major world regions:

- Latin America and the Caribbean
- Sub-Saharan Africa
- The Middle East
- South Asia
- Slavic and Eastern Europe
- Southeast Asia
- East Asia.

## ***2. WNA Phase One: Latin American Newspapers from the Era of Independence, State Formation, and Early Modernization 1805-1923***

Development of the World News Archive (WNA) will unfold in multiple phases, involving creation of discrete but interrelated and interoperable collections of digitized news content. The goal of the initial phase (Phase One) of the endeavor is to digitize and provide continual, persistent Web-based access to an estimated one million pages of Latin American newspapers for CRL, Affiliates, and their respective communities.<sup>1</sup> The titles to be digitized in Phase One (here, “the WNA Content”) are listed in Appendix A. Phase One of the project is expected to be completed on or before December 31, 2010.

It is expected that the methodology and workflows developed in Phase One will constitute a template for the conversion and electronic delivery of additional newspapers from Latin America and other world regions during subsequent phases of the project. CRL envisions that the partnership formed to accomplish Phase One will be evaluated and reaffirmed as CRL undertakes subsequent phases of the World News Archive.

CRL and Affiliates expect to derive from the endeavor benefits, monetary and/or non-monetary, commensurate with their investment in the project.

*To this end, CRL and Affiliates invite electronic publishers, aggregators, and other organizations, both commercial and not-for-profit, (here “Vendors”) to submit proposals for partnership in this endeavor. Specifically, Vendors are invited to propose to provide one or more of the following sets of services:*

- a) Conversion (“Conversion Services”) of the WNA Content from positive microfilm to digital format, and an appropriate level of markup, annotation, OCR conversion, and metadata production.
- b) Persistent and continuous Web-based access to the digitized WNA Content (“Access Services”) for members of designated CRL communities, with minimum functionality specified in Sections 4 and 5 below, and interoperability with related digital newspaper collections and/or other collections of primary source materials relating to Latin America as appropriate.
- c) Other services and benefits of interest to CRL and its communities (“Other Services”). These might include, but are not necessarily limited to, access to additional electronic resources related to the digitized WNA Content, digital conversion of additional collections and materials of interest to CRL and its members, royalties, and archiving of the digitized WNA Content and related metadata to ensure their continued functionality and availability to CRL and its communities.

---

<sup>1</sup> CRL is a consortium of more than 200 North American universities, colleges, and independent research libraries. A list of member institutions may be found at <http://www.crl.edu/content/members/memberListAlpha.asp>.

In their proposals Vendors must indicate the set or sets of services they propose to provide to the endeavor and address the relevant factors enumerated in Sections 4, 5 and/or 6 below. Vendors may propose to provide one or more, or all three, sets of services. Proposals may be submitted by individual organizations or by multiple organizations working in tandem. However, CRL will engage no more than one Vendor for any one of the three sets of services.

If chosen to participate in WNA Phase One, Vendors should expect to invest funds and in-kind services in the endeavor, and can expect to accrue an appropriate return on their investment. The aggregated digital resources developed under the proposed endeavor will be made available exclusively through the partnership, except as indicated in Section 3.3 below, for a specified period of time. Alternatively, CRL might choose to engage a Vendor for Conversion Services only, and to create or obtain from a partner or provider of its own choosing an open or proprietary platform for providing access to the WNA Digital Content.

Notwithstanding the above, CRL and Affiliates will reserve the right to make available to their respective communities the WNA Digital Content derived from Source Materials from their respective holdings in parallel with the WNA, with the overall goal of providing effective widespread access to newspapers through scanning and text conversion and implementing alternative interfaces for navigating and exploring their collections of newspapers.

### **3. General Terms and Conditions**

*3.1 General Conditions:* In this undertaking, CRL and its Affiliates intend to make their holdings of world newspapers (the “Source Materials”) easily accessible in digital format to research and scholarly communities in North America and throughout the world. CRL and Affiliates expect to derive from this endeavor an appropriate return on investment in the form of increased access to research resources for their constituent communities, financial compensation for costs incurred, or some combination of the two.

*Development of the World News Archive will follow the general parameters outlined in this Section. Those parameters represent CRL’s current expectations regarding the nature and scope of the World News Archive endeavor, and constitute neither a commitment to act on the part of CRL, nor a limitation on the eventual actions and activities of CRL in the proposed endeavor.*

The goal of this initial phase (Phase One) of the endeavor is to digitize and present in electronic form an estimated one million pages of Latin American newspapers, from the titles on the attached list. Content will be selected by and drawn from the holdings of CRL and CRL Affiliates listed below.

University of California, Los Angeles  
University of California, Berkeley  
University of Chicago  
University of Florida  
Harvard University  
University of Illinois at Urbana Champaign  
New York Public Library  
University of Texas Libraries  
University of Washington

CRL reserves the right to enlist additional CRL member libraries, as well as libraries from the regions in which the newspapers originate, to participate in this endeavor as Affiliates. CRL also reserves the right to amend the list of titles to be converted (See Appendix A for a list of proposed titles).

*3.2 CRL Obligations and Services:* The Source Materials will be supplied in positive microform format from the collections of CRL and Affiliates. CRL will be responsible for securing for the WNA the film positive copies of the Source Materials for digitization. CRL will provide reasonably complete runs of each newspaper title or, when only incomplete runs are accessible, the entire holdings of a title available to CRL, its Affiliates, and other appropriate partner organizations. CRL will deliver the Source Materials to the Conversion Services Vendor for digitization, and after conversion will store the Source Materials and make them available for later re-conversion and quality assurance purposes if needed.

CRL will be responsible for providing to the Access Services Vendor the WNA Digital Content, associated metadata, and OCR text in the types and formats specified in Section 4.1 below. The schedule and pace of delivery of said materials will be established by agreement between CRL and the Access Services Vendor.

*3.3 Rights and Ownership:* CRL will clear with copyright owners and will secure all rights necessary to permit digital conversion and dissemination of the Source Materials to scholarly users. CRL will provide to the Access Services Vendor a limited license for specified uses of the WNA Digital Content, as appropriate to the Vendor's role in the endeavor, which shall last for a specified duration. The scope and duration of the license will be negotiated between CRL and Access Services Vendor and will be commensurate with the scale of the latter's investment in the endeavor. The license will be renewable periodically by mutual agreement.

CRL will receive escrow copies of all digital files and associated metadata for preservation purposes, and will reserve the right, forward from a specified date to be determined by mutual agreement between CRL and the Access Services Vendor, to make those files available to CRL and its communities. All contracts will stipulate that, after a mutually agreed upon and specified period ("the blackout period"), CRL and Affiliates will thenceforth be free to disseminate and make any and all possible uses of the digital files

and associated metadata created by the Conversion Services Vendor from their respective holdings.

**3.4 *Exclusivity:*** To offer sufficient incentive for participation, the specific rights granted to the Access Services Vendor may be assigned on an exclusive basis. The rights thus granted will in all cases be disclosed to CRL and its Affiliates.

**3.5 *Distribution of Revenues and Benefits:*** The revenues and benefits generated by the digital resources and assets developed under the WNA endeavor shall accrue to CRL and the Access Services Vendor(s) in an equitable manner.

In general, the return on investment to CRL and Affiliates will be proportionate to the value of the content, services and funding provided by CRL and Affiliates. Compensation to CRL might take the form of monetary payments or royalties; digital image files, text files, metadata, translations, and/or other digital products derived from the WNA Content; access to digital resources generated by the endeavor; electronic access to additional digital resources available from the Access Services Vendor; additional services provided by the Vendor(s); and/or other benefits.

**3.6 *Pricing of the Resulting Resource:*** It may be appropriate to provide preferential pricing of the resultant digital resources for certain markets. CRL may set conditions on the Access Services Vendor's pricing of the World News Archive resources that are preferential to certain communities. Specifically, distinct pricing tiers will be established for four markets: CRL Affiliates contributing to the World News Archive, CRL member institutions; libraries and universities in the regions in which the newspaper content originated; and all other users.

**3.7 *Duration of the Agreement:*** The duration of the cooperative agreement(s) between CRL and the Vendor(s) for Phase One will be established by mutual agreement between CRL and the respective Vendor(s).

#### **4.1 *General Requirements for Conversion Services:***

**4.1.1 *Conversion Services Requirements:*** The goal of Phase One of the WNA is to provide continual, persistent Web-based access to non-U.S. newspaper content with sufficient documentation, metadata and functionality to support a wide range of humanities and social science research, teaching and cultural heritage activities and programming.

After the cooperative agreement is executed, CRL will convene a meeting of its Affiliates and the selected Conversion Services and Access Services Vendor(s) to review the technical guidelines, establish work-plan milestones, and, if necessary, further specifications for deliverables. In subsequent phases of the project, the initial approach and associated guidelines will be evaluated and, if appropriate, revised based on stakeholder feedback, experience, and technological advances.

*4.1.2 Provision, Transport and Control of Source Materials:* CRL will be responsible for securing, preparing and delivering to the Conversion Vendor film positive copies (“WNA Source Materials”) for digitization. CRL will provide reasonably complete runs of each newspaper title or, when only incomplete runs are accessible, the entire holdings of a title available to CRL, its affiliates, and partner organizations. An effort will be made to deliver as complete a run of each title as possible. In some cases a limited number of paper copies of the title, or previously scanned images from same, may be substituted to complete the run. CRL will arrange for archival backup of these paper fill-ins on microform as appropriate.

CRL will establish a schedule for delivery of the WNA Source Materials to the Conversion Services Vendor in consultation with said Vendor. The schedule will be contingent upon the availability of the source materials to CRL and the production schedule proposed by the Vendor.

In obtaining microfilm for digital conversion under WNA Phase One, CRL will give preference in its selection to titles on higher quality microfilm, avoiding wherever possible shadows, page curvature, scratches or other disfigurements in microfilm emulsion and base, and other defects. It is understood, however, that due to the vagaries of microfilm production over the past half century, and the inconsistent application of standards, that the image quality of WNA Source Materials will not in every case be high.

CRL will provide a detailed, reel-by-reel listing of WNA Source Materials delivered to Conversion Services Vendor. Each reel shall have a list of the title(s) included thereon, dates of issues of that title included, and, to the extent possible, the dates or ranges of dates of missing or partial issues and any physical defects noted on the reels.

The Conversion Services Vendor will be responsible for ensuring adequate climate control, security, and inventory control of WNA Source Materials while in Vendor custody and while in transit from Vendor to CRL. Vendor is also responsible for the safe and timely return of the WNA Source Materials to CRL, and for indemnifying CRL against any loss of or damage to said materials while same are under the control of Conversion Services Vendor and on return transit. Transport of WNA Source Materials to and from the conversion facility will be through an express carrier acceptable to CRL.

*4.1.3 Conversion Procedures and Requirements:* Conversion Services must preserve the integrity of WNA Source Materials, capturing all physical and structural characteristics of same considered by CRL to be relevant to its communities’ uses. Such characteristics include, at minimum the full layout of each individual page, the integrity of article text within a page and an issue; the original sequence of pages within a given issue; and the original publication sequence of issues within a title.

To achieve this goal CRL technical requirements for Conversion Services for Phase One are designed to balance long-term CRL objectives with near-term constraints. Long-term CRL objectives include:

- Page images of sufficient spatial and tonal resolution to support effective performance of OCR (optical character recognition) software and representation of printed half-tones, given the limitations of microfilm, expecting that future improvements in OCR and image processing will be applied to the same images;
- Digital file formats with a high probability of sustainability, more specifically standard, widely accepted and supported formats;
- Cost-effectiveness of digital conversion and maintenance of the resulting assets.

*4.1.4 Digital Conversion:* The Conversion Services Vendor will create from WNA Source Materials digital images, OCR and text files, associated metadata, and other derivative files as described below. Vendor will deliver to CRL on a schedule established by mutual agreement between the Vendor and CRL, copies of all digitized WNA Content and associated metadata described below (See Appendix B: Technical Guidelines for Conversion and Appendix C: Digital Asset Metadata Elements for more details).

**For each newspaper title**

- Up-to-date MARC record from the CONSER database, fully conformant to current standards for cataloging print newspapers,
- Additional title-level metadata related to the title run/s digitized and delivered.

**For each newspaper issue/edition**

- Structural metadata for issues/editions digitized and organized by date.

**For each newspaper page**

- Page image in two raster formats:
    - 8-bit grayscale, minimally 400 dpi, uncompressed TIFF 6.0,
    - Same image, compressed as JPEG2000,
  - File with OCR text and associated bounding boxes for words or characters, 1 file per page image,
  - PDF Image with Hidden Text with text and image correlated,
  - Structural metadata to relate pages to title, date, and edition, sequence pages within issue or section; to identify image and OCR files; and to support a chronology-based browsing interface and full-text searches with retrieval of individual page images, and highlighting of search terms on said images
- Technical metadata to support the functions of a trusted repository.

Newspapers microfilmed two sheets per frame must be split into two separate image files and assigned appropriate metadata. As grayscale images are requested, de-speckling, edge-enhancement, and similar techniques that are relevant when producing bitonal images are not necessary. To improve appearance and OCR accuracy, images with more than 3 degrees of skew should be de-skewed. Image files should be cropped to the page edge (not to the text block boundaries).

*4.1.5 Metadata:* Each newspaper digitized for the WNA must be supported by coherent metadata, to provide intellectual access and support navigation of the structure of the publication, by date, section, etc. Appendix C lists the elements appropriate at the newspaper title level, the issue/edition level, and the page level. The access interface will permit direct identification and citation at each level through persistent identifiers.

The Conversion Services Vendor should implement the draft standard NISO Z39.87–2006 Data Dictionary – Technical Metadata for Digital Still Images for master images in WNA Digital Content. To support responsible custodianship of these images, the headers for the master TIFF images should incorporate tagged metadata relating to the creation of the images.

*4.1.6 Content Segmentation:* The Conversion Services Vendor may propose to provide either article-level or page-level segmentation in the WNA Digital Content. In either case Vendor should provide a technical plan for achieving the desired results. Page images and related files for each newspaper title shall be organized in a hierarchical directory structure sufficient to permit identification of the individual digital assets from the metadata provided.

*4.1.7 OCR Conversion:* The Conversion Services Vendor must enable full-text searchability of the WNA Digital Content based upon optical character recognition (OCR) or comparable conversion and indexing processing. (If enhancements are integrated into the OCR operation as a pre-processing step, the grayscale image files created and delivered to CRL should have the same enhancements.) See Appendix B: Technical Guidelines for Conversion for more details.

## **4.2 Conversion Services Proposal**

Vendors must address the following elements in their proposals to provide Conversion Services for Phase One of the WNA endeavor. The services proposed here should be those for which the Vendor has demonstrated unique and/or superior capabilities, and a successful track record of delivering. Aside from addressing the specific requirements, the Vendor should offer persuasive evidence of their ability to provide such services more efficiently and at a lower cost than competitors and other potential partner organizations.

*4.2.1 Vendor must detail all services they propose to provide for Phase One of the WNA endeavor; indicate the rate of production they propose to maintain for digital conversion of WNA Content, markup, OCR conversion, and metadata production during the project; and must propose a time frame for accomplishing the Conversion Services.*

*4.2.2 Vendor must describe the climate control, security, and inventory control they propose to provide for WNA Source Materials while in Vendor's custody, and measures that Vendor will take to protect WNA Source Materials while in transit.*

*4.2.3 Vendor must describe how they propose to deal with the following image conversion challenges: high reduction ratio used in microfilming the original newspaper; variations in density within images, or between exposures on a single microfilm reel; master positive of a quality that renders questionable a high level of confidence for OCR.*

*4.2.4 Vendor must specify the form in which they propose to deliver metadata (XML, tab-delimited files, MySQL, or MS Access databases, etc.). Vendor should describe what tags corresponding to mandatory elements from the draft standard NISO Z39.87–2006 they propose to employ, and what additional tags corresponding to recommended elements from Z39.87–2006 and what additional tags necessary for rendering of the images (e.g., tile specifications, if used) will be employed as well.*

*4.2.5 Vendor must describe how OCR will be performed; specifying how OCR text will be correlated to the digital image files and how highlighting of search results text will be achieved. Vendor must indicate how the OCR output elements will be represented in the digital files delivered. This requirement may be met by providing appropriate documentation for the OCR software used. Vendor must provide confidence level data at the page, line, character, and/or word level for their OCR strategy.*

*4.2.6 Vendor must describe in detail what quality assurance procedures and mechanisms will be implemented to ensure the integrity, legibility, interoperability, and usability of the converted files, and to what degree flaws or defects will be tolerated.*

*4.2.7 Vendor must describe the method by which they propose to deliver digital images, OCR and text files, associated metadata, and other derivative files to CRL for backup and archival purposes.*

*4.2.8 Vendor must identify, describe in detail, and justify any proposed deviation from or enhancement to the above specifications and indicate the cost to CRL connected with said deviation or enhancement, if any.*

*4.2.9 Cost: Vendor must identify any and all costs, and/or obligations in addition to the activities that CRL will undertake as described in Section 4.1 above, that Vendor expects CRL to bear in connection with the Conversion Services activities. Vendor must identify the basis of those costs, the factors that may affect costs, and the potential variations in CRL costs that might arise from those cost factors.*

## **5.1 General Requirements for Access Services**

The Access Services Vendor will provide persistent and continuous Web-based access to the digitized WNA Content for members of specified CRL communities, with minimum

functionalities specified in this Section, and interoperability with related digital newspaper collections and/or other collections of Latin American primary source materials as appropriate.

*5.1.1 Minimum Interface Functionalities:* The following minimum functionalities are to be supported by Web-based presentation of the World News Archive Digital Content.

*Browsing:* The interface must allow the user to select the newspaper issue they want to see using a drop-down menu for choosing the newspaper title and a date selection facility, based on a calendar view, to select the individual issue. It must support seamless navigation forward and backward in time between contiguous issues of a given newspaper title.

*Searching:* At minimum the interface must support both simple and advanced searching against the full text of the newspaper pages; the results set returned from a search will present all newspaper pages containing the search term entered and permit direct linking to the page image. It must support the ability for users to search across all or selected newspaper titles. Search terms must be highlighted in the search results and in the PDF page image.

*Viewing:* The interface must present to the user uncropped, right-reading full-page images of each individual page of the newspaper title as it appears in the source materials. The interface should employ a zooming capability based on JPEG2000 wavelet compression. This technology will compress the newspaper image effectively and permit the presentation of image segments dynamically, at the user's request.

*5.1.2 Minimum Browser Configurations:* Access to the WNA Digital Content must be through a Web-based interface that supports viewing through standard Web browsers, including but not limited to the following: Internet Explorer, Internet Explorer for Mac, Netscape, Mozilla Firefox. The interface must also support viewing using the following media players and viewers: Adobe Acrobat or Acrobat Reader.

*5.1.3 Authentication of Users and Tiered Accessibility:* In providing access to the WNA Digital Content Vendor must be capable of distinguishing between, and providing different levels and terms of access to, four distinct CRL communities:

1. CRL Affiliates, who provide Source Materials for the WNA Digital Content;
2. Research libraries and centers in Latin America and the Caribbean, and other regions from which WNA Source Materials originated;
3. CRL members and their local communities, which currently include an estimated 213 research libraries and the authenticated local user populations;
4. All other users.

The user interface provided must be capable of reliably distinguishing between the four communities of users, and of according authorized users the appropriate access rights, services, content, and functionalities as determined by CRL and the Access Services Vendor.

The market(s) to which the Vendor proposes to distribute WNA Digital Content may be all-inclusive, or may be limited to specific sectors or geographic regions. Examples of designated markets are: all markets, the Western Hemisphere, North America, the U.S., the EU, others.

## **5.2 Access Services Proposal**

Vendors must address the following elements in their proposals to provide Access Services for Phase One of the WNA endeavor. The services proposed here should be those for which the Vendor has demonstrated unique and/or superior capabilities, and a successful track record of delivering. Aside from addressing the specific requirements, the Vendor should offer persuasive evidence of their ability to provide such services more efficiently and at a lower cost than competitors and other potential partner organizations.

*5.2.1 The Access Services Vendor must provide a detailed description of the Web-based user interface through which Vendor proposes to provide access to the WNA Digital Content, and must provide access to either a functional prototype or working model of same. The Vendor must identify any and all proprietary software said interface will utilize in enabling search, navigation, and display of WNA Digital Content.*

*5.2.2 Vendor must describe processes to incorporate WNA Digital Content subsequently added to the original content (in the case of fill-in issues or pages or earlier or later dates).*

*5.2.3 Vendor must describe in detail the failsafe and risk mitigation processes, systems, certifications, and mechanisms Vendor has in place to ensure continuous, persistent Web-based access to the WNA Digital Content in the event of interruptions or disruptions to the Vendor's systems and operations.*

*5.2.4 Vendor must describe what technical support by qualified staff will be provided by email and telephone for technical staff at CRL and its community institutions. Vendor should stipulate the hours and days of the week that such technical support is to be available.*

*5.2.5 Vendor must identify related digital content that they propose to make interoperable with WNA Digital Content, and accessible to CRL communities through said interface. Vendor must specify the nature and degree of interoperability between WNA Digital Content and related content, if any, that is to be achieved through the interface. Interoperability might include, for example,*

*cross-search ability in text searches, indexing using a shared controlled vocabulary, etc.*

*5.2.6 Vendor must specify what Web browsers the proposed WNA access interface supports, and what media players and viewers are supported.*

*5.2.7 Vendor must indicate how the WNA access interface will distinguish between and authenticate users in each of the four communities cited above in providing access to the WNA Digital Content.*

*5.2.8 Vendor must specify how they propose to monitor and report to CRL statistical information on usage of the WNA Digital Content by CRL member and Affiliate organizations.*

*5.2.9 Vendor must identify the market(s) in which they propose to distribute the WNA Digital Content and derivative products they propose to produce under the endeavor. Vendor should indicate whether they expect to limit their marketing and distribution to specific segments, such as nonprofit institutions, research libraries and academic institutions, individual users engaged in academic or scholarly research, other.*

*5.2.10 Vendor must identify, describe in detail, and justify any proposed deviation from or enhancement to the requirements in Section 5.1 and indicate the cost to CRL connected with said deviation or enhancement, if any.*

*5.2.11 Vendor must identify and quantify all initial and/or ongoing services, if any, in addition to those specified in Section 3.2 above, that might be required from CRL to enable Vendor to provide the Access Services.*

*5.2.12 Costs: Vendor must identify any and all potential additional costs and/or obligations that Vendor expects CRL and/or its communities to assume, if any, in connection with access to the WNA Digital Content through the Vendor interface. Vendor must identify the basis of those costs and the factors that may affect costs, as well as the potential variations in CRL costs that might arise from those factors.*

*5.2.13 Revenues: Vendor must specify the basis on which any proposed royalties or revenues, if any, derived from use of the WNA Digital Content are to be calculated and accrued. Vendor must specify how frequently statements are to be provided to CRL, and should describe the process in place for the independent verification and auditing of reported revenues and payments. Vendor must specify the timetable upon which royalties or revenues will be disbursed to CRL, e.g., quarterly, semi-annually, annually.*

*5.2.14 Vendor must specify an estimated timeframe for provision of Access Services.*

*5.2.15 Vendor must identify the minimum duration, in years, for which they propose to be granted exclusive rights to host and distribute WNA Digital Content (i.e., the “blackout period”).*

## **6.1 General Requirements for Other Services**

Vendors may propose to provide other services and benefits of interest to CRL and its communities (“Other Services”). These might include, but are not necessarily limited to, access to electronic resources related to the WNA Content, such as related digitized newspaper or news materials or other related primary source materials in digital form. Other Services might also include digital conversion of additional collections and materials of interest to CRL and/or its member and affiliated institutions; archiving of the digitized WNA Content and related metadata to ensure its continued functionality and availability to CRL and its communities; or royalties.

Vendor might also propose a royalty arrangement, whereby monetary payments are made to CRL in return for use of the WNA Content or derivatives and/or related services or assets obtained from CRL.

Finally, Vendors may propose an archiving solution, whereby the Vendor would provide secure hosting of digitized WNA Content and verifiable assurance of their ability and readiness to deliver the WNA Digital Content and associated metadata to CRL on demand.

Vendors might also develop derivative assets, products and/or services through the digital conversion activities, as a way to recoup part or all of their investment in the endeavor.

## **6.2 Other Services Proposal**

Vendors must address the following elements in their proposals to provide Other Services for Phase One of the WNA endeavor. The services proposed here should be those for which the Vendor has demonstrated unique and/or superior capabilities, and a successful track record of delivering. Aside from addressing the specific requirements, the Vendor should offer persuasive evidence of their ability to provide such services more efficiently and at a lower cost than competitors and other potential partner organizations.

*6.2.1 Vendor should identify and describe in detail the related additional electronic resources which they propose to make accessible to CRL and its communities, and indicate how those resources relate to the WNA Digital Content.*

*In instances where said related additional electronic resources are still in development, Vendor must specify when they will come on-line and be available via the Web to CRL and its communities.*

6.2.2 *Vendor must identify the terms and basis upon which they propose to make said additional resources accessible to CRL and its communities, including any costs to CRL and/or its communities, obligations thereto, or limitations upon the use of such additional resources.*

6.2.3 *Vendor should identify and describe in detail additional Latin American newspaper titles that they propose to include in the WNA Digital Content, justifying their value to the WNA corpus and identifying the source of the microfilm copy to be digitized.*

6.2.4 *Vendor should specifically identify any additional collections or source materials from CRL and affiliates' holdings that they propose to convert to digital formats and/or offer through an access interface.*

6.2.5 *Vendor should specify what digital files, associated metadata, and other derivatives will result from this service and will be delivered to CRL for its purposes.*

6.2.6 *Vendor should specify whether they propose to host the additional digital files and metadata and, if so, the terms and basis upon which Vendor will provide access for CRL and its communities.*

6.2.7 *Vendor must indicate clearly what additional rights and licenses they will require CRL to grant to enable Vendor to accomplish conversion of additional collections and source materials and provide associated services.*

6.2.8 *Vendor must specify the basis on which the proposed royalties are to be calculated and accrued.*

*Vendor must specify how frequently statements are to be provided to CRL, and should describe the process in place for the independent verification and auditing of reported revenues and payments.*

*Vendor must specify the timetable upon which royalties will be disbursed to CRL, e.g., quarterly, semi-annually, annually.*

6.2.9 *Vendor must specify what provisions they will put in place to store, maintain, migrate, and otherwise preserve fully functional digital master files, metadata, and enabling software for the WNA Digital Content.*

*Vendor must specify how they propose to obtain independent verification or certification of fulfillment of the aforementioned provisions.*

6.2.10 *Vendor should describe in general the products and services envisioned, if any, and suggest how such products or services might be disseminated, aggregated, and/or otherwise exploited.*

*6.2.11 Vendor must identify any and all costs or obligations that might be incurred by CRL in connection with the proposed Other Services activities, such as the cost of preparation and/or delivery of additional source materials to Vendor, obligation to clear rights with respect to the materials, and so forth. Vendor must identify the basis of such costs, the factors that may affect such costs, and the potential variations in CRL costs that might arise from those cost factors.*

*6.2.12 Vendor must specify an estimated timeframe for provision of Other Services.*

## **7.1 Submission and Evaluation of Proposals**

This section sets out the general terms and conditions under which CRL is prepared to contract for services under this endeavor. If a Vendor is not prepared to comply with such terms, in whole or in part, it must set out its alternative position, in detail, in a paragraph-by-paragraph response. Failure to set out an alternative position will be construed as unconditional agreement with CRL's stated contractual terms.

*7.1.1 Evaluation of Proposals:* In choosing a Vendor or Vendors, CRL will favor organizations that bring to the endeavor the greatest combination of competency and experience in the conversion and distribution of digital resources, and commitment of in-kind and financial investment to the project. CRL will also view favorably Vendors who submitted proposals to CRL under the World News Archive RFI issued in 2006.

Proposals will be evaluated by CRL in terms of the following weighted criteria:

Value versus cost to CRL of services and benefits proposed	<b>45%</b>
Technical Soundness of Vendor Proposal	<b>30%</b>
Demonstrated Capability of Vendor to Provide Services	<b>15%</b>
Vendor proposal submitted under the 2006 RFI	<b>10%</b>
<b>Total:</b>	<b>100.00%</b>

*7.2.2 Proposal Deadline, Delivery and Copies Required:* Vendors wishing to respond to the RFP shall submit to CRL two (2) complete copies of their Proposal, plus one electronic (CD, DVD, or email attachment) copy of the Proposal, in Microsoft Word format, in a sealed container by the closing date and time stated below. Proposals shall be marked:

CRL World News Archive Project  
RFP Number 200701

*7.2.3 Due Date:* 4:00 p.m., Central Daylight Time (CDT), May 15, 2007.  
The container shall be addressed to:

Center for Research Libraries  
6050 South Kenwood Avenue  
Chicago, IL 60637  
Attention: James Simon

Each Vendor shall assume full responsibility for delivery and deposit of the completed proposal no later than the specified due date above. CRL is not responsible for any loss or delay with respect to the delivery of the proposals. **ANY PROPOSALS RECEIVED AFTER THE DATE AND TIME SPECIFIED WILL NOT BE CONSIDERED, AND WILL BE RETURNED UNOPENED.**

*7.2.4 Announcement of Proposals:* CRL will announce the names and addresses of Vendors who submitted proposals, by 10:00 a.m. CDT on May 18, 2007. No announcement will be made regarding the prices or other information contained in the proposals.

*7.2.5 Proposal Content and Format:* Proposals must be succinct documents that directly address the requirements of Sections 4, 5, and/or 6 of this RFP. Proposals that do not address all of the requirements of the appropriate Section shall be deemed **NONCOMPLIANT** and shall be ranked in a separate group after all compliant proposals.

*7.2.6 Firm Prices:*

a) All cost estimates provided in the Proposal will be considered final for the purposes of the evaluation process. Actual pricing for services under this contract will be based upon an agreed pricing methodology negotiated with successful Vendors that will in no case be greater than the costs estimated in the Vendors' proposals.

b) Vendors are responsible for including all costs and shall use their prior experience to ensure that all costs (including costs for all items not specifically identified in the RFP) are included. The proposed cost estimates submitted in the proposal will be binding on Vendors and available for CRL to accept as proposed.

*7.2.7 Completeness of Proposals:* Vendors are solely responsible for the completeness and accuracy of their proposals and of the answers to CRL requests for clarification. CRL accepts no responsibility for any incompleteness or inaccuracy of any proposal or clarification no matter how caused.

*7.2.8 Proposal Validity Period:* The proposal remains valid for 120 days following the closing date of the RFP.

*7.2.9 Withdrawal of Request for Proposal:* CRL may withdraw this RFP at any time and shall not be liable for any expenses, costs, loss or damage incurred or suffered by any Vendor as a result of such withdrawal.

*7.2.10 Acceptance or Rejection of Proposals:* Notwithstanding any other provision in this RFP, CRL can, at its sole discretion: accept any Proposal; accept parts of more than one Proposal; reject any Proposal; and reject all Proposals.

*7.2.11 Vendor's Costs and Expenses:* Any and all costs incurred in responding to this RFP, responding to the CRL questions, conducting demonstrations or any other related activities shall be borne by the Vendor. The rejection of any or all proposals shall not render CRL liable for any costs, expenses or damages incurred by any Vendor. All proposal documents will become the property of CRL.

*7.2.12 Query Period:* There will be a Query Period following release of the RFP, ending two weeks prior to closing during which time Vendors may request clarifications in writing (see below). Questions received after the end of the Query Period may not receive a response. Vendors' questions and the answers they receive will not be communicated to other Vendors unless they result in a clarification that, in the opinion of CRL, would result in an unfair advantage to the Vendor posing the question in which case such clarifications will be published by addendum.

*7.2.13 Timetable:* This RFP process will follow this timetable:

RFP Release March 20, 2007  
RFP Query Period March 21-May 1, 2007  
RFP Close May 15, 2007  
Evaluation Period May 16-June 14, 2007  
Contract Negotiations Commence June 15, 2007

CRL intends to proceed expeditiously but reserves the sole right to modify dates in the above timetable.

*7.2.14 RFP Interpretations and Addenda:* Should Vendors discover discrepancies, ambiguities, or omissions in this RFP document, or should Vendors be in doubt as to the meaning of any requirement or process in this RFP, they should immediately notify:

James Simon, Director of International Resources  
Center for Research Libraries  
6050 South Kenwood Avenue  
Chicago, IL 60637  
USA  
Bus: (773) 955-4545, ext 324  
Bus Fax: (773) 955-4339  
Email: [simon@crl.edu](mailto:simon@crl.edu)

Vendors are solely responsible for identifying and seeking such clarifications to the specifications as may be required to fully understand and cost the work to be carried out in the event of contract award. CRL will not consider any claims for additional payments not identified in the Contract. Questions about the intent or meaning of any part of this RFP

may be asked during the Query Period. Questions must be in writing to avoid any misunderstandings and may be sent by facsimile or email. CRL will not be responsible or liable for answering verbal inquiries or for the accuracy of verbal answers. CRL will endeavor to respond in writing within three (3) days of receiving the question.

All interpretations and any supplemental instructions provided by CRL will be in the form of written addenda, which shall be deemed to be given at time of delivery or facsimile transmission, or three (3) days following the date of mailing, whichever is less. Such addenda shall become part of the RFP document and, in its form of proposal; Vendors shall acknowledge receipt of all addenda by listing each addendum and its date of issue.

*7.2.15 Additional Information:* CRL may request additional information or clarification about a Vendor's proposal by sending a written inquiry to an address that the Bidder shall provide in its Proposal. Vendors shall provide a timely written response, normally within five (5) working days of receipt of the inquiry and such requests and responses shall form part of the proposal.

*7.2.16 Conflicts of Interest:* Vendors are required to disclose to the CRL Director of International Resources, prior to the ranking of the Vendors by CRL, any potential conflict of interest. A potential conflict of interest includes influential personal relationships between the CRL, its officials, officers, employees, agents or independent consultants and Vendors, its officers, employees, agents, subsidiaries and parent organizations. If a conflict of interest appears to exist as referred to above, the CRL Director of International Resources may, at his discretion, withhold the ranking of the Vendors until the matter is resolved to his satisfaction and, if necessary, to the satisfaction of the appropriate CRL authorities.

*7.2.17 Disclosures:* Although all proposals from prospective partner organizations will be considered confidential, the endeavor will be characterized by transparency and the open sharing of the terms of the agreements and contracts between CRL and partner organizations with CRL stakeholders to the extent possible.

## Appendix A: List of Newspaper Titles for Phase One

The titles listed below represent a subset of the holdings of Latin American and Caribbean newspapers held by CRL and Affiliates.

<b>La Nacion (Buenos Aires, Argentina)</b>	[1906-1923]	104 reels	57,200 pages
<b>La Prensa (Buenos Aires, Argentina)</b>	[1857; 1869-1923]	245 reels	134,750 pages
<b>La Razon (Buenos Aires, Argentina)</b>	[1920-1923]	12 reels	6,060 pages
<b>Review of the River Plate (Buenos Aires, Argentina)</b>	[1919-1923]	5 reels	2,750 pages
<b>Standard (Buenos Aires, Argentina)</b>	[1861-1881]	27 reels	13,365 pages
<b>Vanguardia (Buenos Aires, Argentina)</b>	[1894-1923]	67 reels	36,850 pages
<b>Correio da Manha (Rio de Janeiro, Brazil)</b>	[1903-1920]	12 reels	6,600 pages
<b>O Estado de Sao Paulo (Sao Paulo, Brazil)</b>	[1875-1923]	129 reels	70,950 pages
<b>Jornal do Comercio (Rio de Janeiro)</b>	[1845-1912]	106 reels	58,300 pages
<b>Mercurio (Santiago, Chile)</b>	[1914-1923]	61 reels	33,550 pages
<b>La Prensa (Havana, Cuba)</b>	[1913-1920]	16 reels	8,250 pages
<b>Diario de Centro America (Guatemala, Guatemala)</b>	[1873-1923]	44 reels	24,200 pages
<b>Daily Chronicle (Georgetown, Guyana)</b>	[1881-1897]	44 reels	24,200 pages
<b>Diario (Mexico City, Mexico)</b>	[1907-1914]	18 reels	9,900 pages
<b>Diario de Mexico (Mexico City, Mexico)</b>	[1805-1817]	9 reels	4,950 pages
<b>El Dictamen (Veracruz Llave, Mexico)</b>	[1902-1923]	~50 reels	27,500 pages
<b>Excelsior (Mexico City, Mexico)</b>	[1917-1923]	30 reels	16,500 pages
<b>Imparcial (Mexico City, Mexico)</b>	[1896-1914]	26 reels	14,300 pages
<b>Mexican Herald (Mexico City, Mexico)</b>	[1895-1915]	73 reels	40,150 pages
<b>El Monitor Republicano (Mexico City, Mexico)</b>	[1846-1895]	41 reels?	22,550 pages
<b>Nueva Era (Mexico City, Mexico)</b>	[1911-1913]	8 reels	4,400 pages
<b>El Pais (Mexico City, Mexico)</b>	[1899-1914]	43 reels	23,650 pages
<b>La Patria (Mexico City, Mexico)</b>	[1855-1913]	8 reels	4,440 pages
<b>La Revista de Yucatan (Merida, Mexico)</b>	[1913-1923]	5 reels	2,750 pages
<b>El Siglo Diez y Nueve (Mexico City, Mexico)</b>	[1841-1896]	60 reels	33,000 pages
<b>El Tiempo (Mexico City, Mexico)</b>	[1891-1900]	3 reels	1,650 pages
<b>Two Republics (Mexico City, Mexico)</b>	[1867-1900]	26 reels	14,300 pages
<b>Universal (Mexico City, Mexico)</b>	[1848-1855]	12 reels?	6,600 pages
<b>Voz de Mexico (Mexico City, Mexico)</b>	[1879-1909]	37 reels	20,350 pages
<b>Estrella de Panama (Panama, Panama)</b>	[1858-68; 1872-1904; 1876-1907]	~60 reels?	33,000 pages
<b>Star &amp; Herald (Panama, Panama)</b>	[1849-1914]	135 reels	74,250 pages
<b>El Comercio (Lima, Peru)</b>	[1839-1911]	38 reels?	20,900 pages
<b>El Peruano (Lima, Peru)</b>	[1827-1923]	95 reels	52,250 pages
<b>West Coast Leader (Lima, Peru)</b>	[1912-1940]	23 reels	12,650 pages
<b>Port of Spain Gazette (Port of Spain, Trinidad)</b>	[1869-1906]	41 reels	22,550 pages
<b>Venezuelan Herald (Caracas, Venezuela)</b>	[1898-1904]	2 reels?	1,100 pages

1,715 reels 940,715 pages

## **Appendix B: Technical Guidelines for Conversion**

Digital images and metadata shall conform to the technical guidelines in this document. Where proposed vendor services differ from these guidelines, the Vendor shall provide justification of any deviation of or enhancement to the guidelines.

In general, all specifications shall be minimally compatible with technical specifications of the National Digital Newspaper Program (NDNP). Technical specifications are available at: <http://www.loc.gov/ndnp/techspecs.html>

### **Requirements for Scanning & Master Images**

1. Spatial resolution shall be minimally 400 pixels-per-inch (ppi) relative to the original newspaper.
2. 8-bit grayscale
3. TIFF 6.0 uncompressed
4. Newspapers microfilmed two sheets per frame must be split into two separate image files and assigned appropriate metadata.
5. De-skew images with more than 3 degrees of skew
6. Image files should be cropped to the page edge (not to the text block boundaries).
7. Capture microfilm target frames. These image files to be identified as “targets” in metadata; will not be used for display.
8. Capture scanning resolution targets at the start of each session, to monitor scan quality. These target images should be delivered with microfilm targets and page images.
9. TIFF headers shall incorporate data as required by the TIFF 6.0 specification, with additional clarifications or requirements as needed and agreed upon prior to commencement.

All operations that change the image dimensions, spatial resolution, or orientation (e.g., cropping, deskewing) must be made to the TIFF before OCR, since the OCR output is expected to include bounding-box coordinates to relate words and characters to their position on the page in the search interface. The grayscale master TIFF files delivered to CRL must have the same characteristics with respect to cropping, deskewing, or other enhancements as the images used for OCR,

### **Requirements for OCR**

1. One OCR text file per page image. (Discrete text files should be produced for each page, rather than for a multi-page issue or entire title).
2. Each OCR text file name will correspond to the page image it represents.
3. Text will be in ASCII or UTF-8 character set.
4. No graphic elements will be saved with the OCR text;

5. OCR text will reflect the text arrangement of the original newspaper and will be ordered column-by-column (that is, in a natural reading order);
6. OCR text files with bounding-box coordinate data will be at the character or word level.
7. OCR software documentation for the representation of all elements listed here (e.g., the bounding-box coordinates) provided, showing how the elements are represented in the actual digital files.

The page images delivered to CRL must correspond in dimensions, orientation, and skew to those used for the OCR. Once a page image is processed by the OCR application, the resulting text must not be changed.

### **Requirements for derivative images**

#### JPEG 2000

1. One JPEG2000 image for each page image.
2. File will be prepared after any image processing is performed.
3. File will correspond with the image that is used for OCR.
4. Height and width will be the same as the TIFF master file.
5. Compressed so that it is about one-eighth of the TIFF or 1 bit per pixel.
6. Each JPEG2000 image will incorporate appropriate metadata to allow users of downloaded images to identify the source publication, date and page number, reel and sequence number of each image.

#### PDF

1. One PDF Image with Hidden Text for each page image.
2. Each searchable PDF file name corresponds to the page image it represents.
3. The PDF files should incorporate adequate metadata to allow users of downloaded images to identify the source publication, date and page number, reel and sequence number of each image.
4. The PDF will not contain any bookmarks, links, named destinations, comments, forms, Javascript actions, external cross references, alternate images, embedded thumbnails, annotations, or private data.

## **Appendix C: Digital Asset Metadata Elements**

CRL and Affiliates shall review and approve a set of core metadata elements for use in the World News Archive initiative. For the purposes of this RFP, metadata specifications shall conform to the “Digital Asset Metadata Elements” specifications of the National Digital Newspaper Program (NDNP). Technical specifications are available at:  
<http://www.loc.gov/ndnp/techspecs.html>