

**REQUEST FOR INFORMATION (RFI)  
FOR USPTO's PATENT DOCUMENT CONVERSION AND XML GENERATION SOLUTION  
DOC-52-PAPT-1200049**

**THIS IS A REQUEST FOR INFORMATION (RFI) BEING RELEASED PURSUANT TO FEDERAL ACQUISITION REGULATION (FAR) PART 10: MARKET RESEARCH.** This RFI is issued solely for informational, market research, and planning purposes only. It does not constitute a Request for Proposal (RFP) or a promise to issue an RFP in the future. This RFI does not commit the Government to contract for any supply or service whatsoever. Further, the United States Patent and Trademark Office (USPTO) is not at this time seeking proposals, and will not accept unsolicited proposals. Respondents are advised that the United States (U.S) Government will not pay for any information or administrative cost incurred in response to this RFI. All costs associated with responding to this RFI will be solely at the responding party's expense. Responses to the RFI will not be returned. Please be advised that all submissions become Government property and will not be returned. Not responding to this RFI does not preclude participation in any future RFP, if any is issued. Responses to this notice are not offers and cannot be accepted by the U.S Government to form a binding contract. It is the responsibility of the interested parties to monitor the Federal Business Opportunities ([www.fbo.gov](http://www.fbo.gov)) site for additional information pertaining to this RFI.

## **1.0 RFI OBJECTIVE**

The USPTO is in the process of transforming the patent application process from an electronic image based system that uses TIFF images, to a text based system based on a custom XML format. Patents End to End (PE2E) is the United States Patent & Trademark Office's (USPTO) project to deliver a next-generation IT infrastructure supporting Patents business operations. This project will replace the nearly four dozen aging legacy systems used today with a single system that unifies electronic processing over the entire patent application lifecycle (hence "end-to-end"). Most relevant to this RFI, the new system will replace documents that are currently represented as scanned TIFF images with documents represented as structured text in XML.

This RFI seeks to obtain information from interested parties, including the vendor community, about potential solutions to the problem of converting image based documents to XML format with an automated solution. The Office is exploring a wide range of possible solutions for several different but related needs.

The aim of this market research is to find solutions for the continuous automated capture and conversion of three patent application document types for every in-coming patent application filing and follow-on submission. These document types include the specification (SPEC), claims (CLM), & abstract (ABST) from patent application documents. These documents must be converted to text and transformed into a custom XML format. Over time, the solution would be adaptable to allow for additional document types such as the information disclosure statement (IDS).

**REQUEST FOR INFORMATION (RFI)  
FOR USPTO's PATENT DOCUMENT CONVERSION AND XML GENERATION SOLUTION  
DOC-52-PAPT-1200049**

**1.1 Problem Scope**

The ideal solution has the following:

Feature #	Description
1.1	Convert application text (i.e. specification, claim, and abstract) from application documents to XML for every in-coming filing and follow-on document
1.2	Provide XML files to the USPTO within 2 business hours of receiving the application documents
1.3	Process 750,000 applications per year with a growth rate of upto 10% per year; each application will have 3 documents with a page count average between 50-70 pages per application
1.4.1	Each extracted application provided shall be individually zipped—one zip file for each application serial number in the list of specified applications
1.4.2	<p>The extracted application (i.e. 07/289,355) shall fall under the following directory structure:</p> <pre style="margin-left: 40px;"> -07    --28      -93        ---07289355.zip</pre> <p>and the contents of the zip file shall follow the structure below:</p> <pre style="margin-left: 40px;"> -- DWY6N6JKPPOPPY2      -- 00000001.tif  -- DWZCGPD3PPOPPY2      -- 00000001.tif  -- DWZCGPE2PPOPPY2      -- 00000001.tif      -- 00000002.tif      -- 00000003.tif      -- 00000004.tif      -- 00000005.tif      -- 00000006.tif      -- 00000007.tif      -- 00000008.tif      -- 00000009.tif      -- 00000010.tif      -- 00000011.tif      -- 00000012.tif  -- DWZCWVQAPPOPPY2      -- 00000001.tif      -- 00000002.tif  -- errors.xml  -- metadata.xml</pre> <p>10 directories, 1060 files</p>

**REQUEST FOR INFORMATION (RFI)  
FOR USPTO's PATENT DOCUMENT CONVERSION AND XML GENERATION SOLUTION  
DOC-52-PAPT-1200049**

Feature #	Description
1.4.2.1	Each directory (e.g. DWY6N6JKPPOPPY2) inside the .zip shall represent an individual document
1.4.2.2	Each file (e.g. 00000001.tif under the DWY6N6JKPPOPPY2 directory) under a directory shall represent one page of the document
1.4.2.3	<p>The "metadata.xml" file shall contain the additional metadata (date, doc code, etc.) necessary for deriving the final naming convention from the contents of the .zip. For application, 11/891,294, the metadata.xml file is shown below:</p> <pre style="text-align: center;"> &lt;?xml version="1.0" encoding="UTF-8"?&gt; &lt;PatentApplication xsi:schemaLocation="urn:us:gov:doc:uspto:patent Untitled1.xsd" xmlns="urn:us:gov:doc:uspto:patent" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"&gt;   &lt;ApplicationNumber&gt;11891294&lt;/ApplicationNumber&gt;   &lt;TechnologyCenterCode&gt;1600&lt;/TechnologyCenterCode&gt;   &lt;Document&gt;     &lt;MailRoomDate&gt;08/09/2007&lt;/MailRoomDate&gt;     &lt;DocumentCode&gt;TRNA&lt;/DocumentCode&gt;     &lt;DocumentKeyID&gt;F5E4QTDIPPOPPY5&lt;/DocumentKeyID&gt;     &lt;TotalPageQuantity&gt;4&lt;/TotalPageQuantity&gt;   &lt;/Document&gt;   &lt;Document&gt;     &lt;MailRoomDate&gt;08/09/2007&lt;/MailRoomDate&gt;     &lt;DocumentCode&gt;SPEC&lt;/DocumentCode&gt;     &lt;DocumentKeyID&gt;F5E4QTDYPPOPPY5&lt;/DocumentKeyID&gt;     &lt;TotalPageQuantity&gt;28&lt;/TotalPageQuantity&gt;   &lt;/Document&gt; &lt;/PatentApplication&gt; </pre>
1.4.3	Extracted applications shall be transmitted via secure high-speed data line
1.5.1	Prior to OCR, convert single page TIFF images to multi-page TIFF images
1.5.2	Prior to OCR, strip certain artifacts (i.e. formulas, equations, tables) in the application document image and convert to a format (i.e. SVGZ or XML) usable by a computer. (USPTO can provide this utility if necessary)

**REQUEST FOR INFORMATION (RFI)  
FOR USPTO's PATENT DOCUMENT CONVERSION AND XML GENERATION SOLUTION  
DOC-52-PAPT-1200049**

Feature #	Description
1.6	<p>Perform OCR using PrimeOCR version 5.3. During OCR, the following PrimeOCR configuration settings apply:</p> <ul style="list-style-type: none"> <li>• Version: 5.3 (or latest version)</li> <li>• Number of Voting Engines: 6</li> <li>• Input Format: TIFF</li> <li>• Input Format Margins: ¼"</li> <li>• Output Format: XML-PRO</li> <li>• Pages: Process All</li> <li>• Processing: Deskew, Despeckle</li> <li>• Language: US English</li> <li>• Autozone: No Autozone (Full Zone)</li> <li>• AutoRotation: No AutoRotate</li> </ul>
1.7	Transform OCR output (XML-PRO) to USPTO XML format and recombine with stripped and converted artifacts.
1.8	Perform Quality Control review of processed files
1.9.1	After processing and review, transmit the XML files back to the USPTO securely via high speed data line
1.9.2	<p>The files returning to the USPTO shall have a top-level folder, PRIME and follow a directory structure similar to the one shown below for application number (12/716,336):</p> <pre> PRIME (PRIME OCR XML FILES top level folder)  - 12     - 71        - 63           - 12716336              ----12716336.2010-03-03.G6CAG7UTPPOPPY2.ABST.xml              ----12716336.2010-03-03.G6CAG7VNPPOPPY2.CLM.xml              ----12716336.2010-03-03.G6CAG816PPOPPY2.SPEC.xml </pre> <p>and under the top-level folder, the application folders shall have compression applied.</p>

**REQUEST FOR INFORMATION (RFI)  
FOR USPTO's PATENT DOCUMENT CONVERSION AND XML GENERATION SOLUTION  
DOC-52-PAPT-1200049**

Feature #	Description
1.9.3.1	<p>Within each application folder, the filename of the XML file shall consist of four fields separated by periods (extra white space has been inserted below to make the fields clear, this whitespace forms no part of the specification):</p> <pre style="text-align: center;"> 12716336 . 2010-03-03 . G6CAG816PPOPPY2 . SPEC .xml ↑ F1          ↑ F2          ↑ F3          ↑ F4 </pre> <p>The definition of each field is as follows:</p> <p>F1: Case serial number, 9 numeric digits, left side zero filled.  F2: Document filing date, ISO format YYYY-MM-DD  F3: IFW DMS key (IFW unique database ID for the document)  F4: Doc Code of document from IFW.</p>
1.9.3.2	<p>For future expansion purposes, in field 4 (F4), any characters not in the set [1-9a-zA-Z] shall be encoded as +xx (a plus sign followed by two hexadecimal characters). Where xx is the hex ASCII value of the character. This covers doc codes that contain dots, or slashes, or other interesting punctuation characters (e.g. "A..." for amendments).</p>

## 1.2 SECURITY

The solution must comply with the minimum controls required by Department of Commerce in accordance with NIST 800-53A Rev 1 and Rev 3. NIST 800-37 Rev 1, NIST 800-53 Rev 3 pertain, and are available for review by prospective vendors at: <http://csrc.nist.gov/publications/PubsSPs.html>.

## 2.0 ACQUISITION STRATEGY

The USPTO acquisition strategy alternatives are still under development. The acquisition strategy will be partially dependent upon the solutions offered as a result of this RFI.

The respondent is encouraged to identify and provide any unique solutions that will result in effective/efficient operations.

### 2.1 Inquiries

Those who wish to submit questions concerning the RFI may do so by e-mail to the Contracting Officer, [VAnne.Tugbang@uspto.gov](mailto:VAnne.Tugbang@uspto.gov) by Monday, March 26, 2012, 10:00 a.m. Contracting Officer's local time. Please include the RFI number in the subject line. All questions and answers will be made available via a modification to the RFI posted on

**REQUEST FOR INFORMATION (RFI)  
FOR USPTO's PATENT DOCUMENT CONVERSION AND XML GENERATION SOLUTION  
DOC-52-PAPT-1200049**

the Federal Business Opportunities website ([www.fbo.gov](http://www.fbo.gov)) no later than 1 week before the due date of the RFI responses. Those parties that are interested in this project will be requested to provide written responses which discuss their technical solutions and the feasibility of their approach. The written responses may include other alternatives and solutions for USPTO consideration. Written responses will not be returned and become the property of the USPTO.

## **2.2 Instructions for RFI Responses**

Responses must be submitted electronically to the e-mail address below:

U.S. Patent and Trademark Office  
Office of Procurement  
ATTN: V'Anne Tugbang, Contracting Officer

EMAIL: [VAnne.Tugbang@uspto.gov](mailto:VAnne.Tugbang@uspto.gov)

Respondents must submit their responses via email in pdf format no later than **Monday, April 9, 2012, 10:00 a.m. Contracting Officer's local time.** The response should be no more than 15 pages in length and no larger than 5 megabytes and should use font size 12 or larger.

Proprietary information submitted in response to this RFI will be protected from unauthorized disclosure as required by the Federal Acquisition Regulation (FAR). All proprietary markings should be clearly delineated. The respondent shall identify where data is restricted by proprietary or other rights and mark it accordingly.

The format for the RFI responses is described below:

The cover page shall contain (1) Company name, (2) Primary Point of Contact, (3) Phone Number and Email Address, (4) Cage Code, (5) NAICS Code, (6) Business Size, and (7) Federal Supply Schedule (FSS) Contract Number, if applicable.

**Introduction:** Provide a brief description of existing capability to perform the requirements or provide proposed Statement of Work language for the services and/or any proposed solution. In the event your company chooses to provide information subject to inclusion in a future RFP Statement of Work (SOW), clearly identify those portions and provide any appropriate authorizations for release of that portion of information within any subsequent RFP SOW issued by the USPTO, exclusive of any proprietary markings.

**Technical Capability:** The respondent's technical ability shall describe the services and/or any product solution(s) or dataset for the areas described in Paragraph 1.0 and 1.1 of this RFI. The responses should include an overall description of the proposed services and/or any product solution(s) and provide technical data and a demonstrated ability for

**REQUEST FOR INFORMATION (RFI)  
FOR USPTO's PATENT DOCUMENT CONVERSION AND XML GENERATION SOLUTION  
DOC-52-PAPT-1200049**

those areas identified. The descriptions should include schedule information for delivery of services and/or product(s); and the technical rationale for providing these to the USPTO. Interested parties should provide information on their ability to use existing assets or procure, customize/configure, maintain and/or provide technical support for the resources needed for the proposed services and/or product(s). Interested parties should also describe technical benefits of their proposed services and/or product solution(s) in terms of existing technologies or resources, improvements/enhancements, cost efficiencies of their specific approach, and any other support capabilities that provide service and/or product excellence or uniqueness.

**Organization Experience/Past Performance:** Provide a brief description of your organization's experience in same or similar services and/or product solution(s) to both commercial and government organizations; and optionally up to three references for same or similar services and/or any product solution(s) should be provided.

Not responding to the RFI does not preclude participation in any future RFP. If a solicitation is released, it will be issued via the Federal Business Opportunities website ([www.fbo.gov](http://www.fbo.gov)). It is the responsibility of the potential offerors to monitor this website for any information that may pertain to this RFI or a future RFP. The information provided in this RFI and any future changes to the RFI are subject to change and are not binding on the USPTO.

Participation in this effort is strictly voluntary. All costs associated with responding to this RFI will be solely at the interested respondent's expense. Respondents are advised that the United States (U.S) Government will not pay for any information or administrative cost incurred in response to this RFI. The objective of this RFI is to assess vendor capabilities and interest. Review of the responses to the RFI will focus on the offeror's technical capability to provide a quality solution, corporate experience/past performance for same or similar activity with commercial activities or government agencies, and responsiveness to the RFI.

### **2.3 RFI Response Due Date**

Please submit information via e-mail to V'Anne Tugbang, Contracting Officer, at [VAnne.Tugbang@uspto.gov](mailto:VAnne.Tugbang@uspto.gov) no later than April 9, 2012 10:00 a.m. Contracting Officer's local time.

### **2.4 RFI Response Contact**

Respondents to this RFI shall designate a primary and one alternate point of contact within the company (Name, Address, Email, and Telephone Number).

### **2.5 Clarification of RFI Responses**

To fully comprehend the information contained within a response to this RFI, there may be a need to seek further clarification from those respondent(s) identified as capable.

**REQUEST FOR INFORMATION (RFI)**  
**FOR USPTO's PATENT DOCUMENT CONVERSION AND XML GENERATION SOLUTION**  
**DOC-52-PAPT-1200049**

This clarification may be requested in the form of brief verbal communication by telephone; written communication; electronic communication; or a request for a presentation of the response to a specific USPTO group or groups. The USPTO reserves the right to seek additional information from those vendors identified with unique solutions that are determined to be beneficial to the USPTO.