REED TECHNOLOGY
*Transforming Data, Delivering Information™*

# Reed Technology and Information Services Inc.

**7 Walnut Grove Drive,**
**Horsham, PA 19044**
**215 441-6400**

Response to Request for Proposal

from

## The U.S. Patent and Trademark Office

for

## Public Dissemination of Data

## DOC50PAPT1300005

## January 22, 2013

Submission of this response indicates that Reed Technology and Information Services Inc. agrees fully with all terms, conditions and provision included in the solicitation and agrees to furnish any or all items upon which prices are offered at the price set opposite each item.

# Volume I

## 1 Introduction

Reed Technology and Information Services Inc. (Reed Tech) is pleased to submit this response to the USPTO's RFP for Public Dissemination of Data.

Reed Tech has the capacity, domain expertise and technical capabilities to perform the services envisioned in the RFP. We have a very broad and deep understanding of the vast majority of the data that the USPTO is seeking to disseminate, and currently process much of it on a continuous basis, creating publication deliverables for the Agency. The work we do for the USPTO, including the magnitude of the U.S. patent data that we process, is indicative of our capabilities and experience and is relevant to our ability to serve the Agency. Services provided to the USPTO include conversion of patent application documents into various digital formats and production of many of the USPTO's publicly disseminated products. In an operation that continuously manages a pipeline with a large and fluctuating volume of all non-secret U.S. patents and applications, 52 weeks a year, typical yearly throughput volumes are:

- 600,000+ granted patents and patent applications processed
- 170 million+ source pages reviewed (EFS and paper)
- 36 million+ source pages processed with OCR (14 million+ processed in Front End Processing and 22 million+ pages in the Grant and Pre-Grant publication processes)
- 36 billion+ characters processed into XML-tagged data
- 5 million + drawings cleaned up and de-skewed
- 12 million+ composed pages produced for publication
- Approximately 1 terabyte of data delivered to the USPTO

In our work with the USPTO, Reed Tech performs data standardization for every incoming patent application flowing through the U.S. system, producing publication deliverables that routinely exceed a 99.995% level of quality (effectively achieving fewer than five character errors every 100,000 characters). This standard has been consistently met or exceeded throughout the more than 40 years we have been working with the Agency in preparing more than half a million pages of patent content for publication on a weekly basis. In support of our work for the USPTO, we continuously maintain over [redacted] of patent content in our data center in Horsham, PA.

Reed Tech is part of LexisNexis®, a global information business and a major aggregator of intellectual property content that serves the legal, corporate and government markets around the world. Among its products is TotalPatent™, which provides access to an over 40-terabyte data store of global patent information, including:

- 100 patent authorities, of which 31 are searchable in English
- Over 65 million PDFs of global patent documents, compressed to one-third of their original size
- Over 56 million patent families
- Over 35 million full-text English translations of foreign patent documents
- A wide range of additional patent content

Reed Tech is therefore well positioned with the requisite capabilities needed to meet the requirements of the USPTO's Public Dissemination of Data RFP.

# 2 Performance Work Statement

This section is intended to address the requirement of providing a Performance Work Statement as found in Section L.6 of the RFP. Associated resources are found in Section 3 of this document.

## 2.1 Technical approach

### 2.1.1 Background

The technical approach has been broken down into three main components: data acquisition, the data repository and public access. Please refer to the following diagram as each of these is discussed.

### 2.1.2 Data acquisition

#### 2.1.2.1 USPTO-Prepared Bulk Data

The bulk data transfer process will download the appropriate patent and trademark data via secure high-speed data line. The historical bulk data will be downloaded on an agreed schedule at the beginning of the contract. Optionally, this historical data could be provided by the USPTO using storage media.

Ongoing bulk datasets will be downloaded based on the individual product's schedule. As the bulk data is loaded into the repository, database records will be updated to make the archives dynamically available on the appropriate website download page. The data will be loaded to the repository in unaltered form and made available to the public free of charge. Any updated files made available by the USPTO (due to missing or corrupt files) will be added or replaced in the repository.

#### 2.1.2.2 Public PAIR Data

**(Reference RFP C.3.2)**

Public PAIR data must be compiled by crawling the USPTO site, that is, using an index or other method to build appropriate web page URL's, load the html and scrape the appropriate data. Scraping is accomplished by utilizing intelligent technology to find the proper data sections located within a complex web page, for example, locating the meta-data fields within the Public PAIR application web pages and downloading the appropriate data.

The desire is for the repository to be completely up to date at the time the dissemination site is first made available to the public. [redacted]

*2.1.2.2.1* [redacted]

[redacted] will access USPTO Public PAIR during the scheduled access times (daily four-hour window) and download new and updated application records. Newly published patent applications can be identified using the weekly Patent Application Publication Bibliographic data.

In order to keep the dataset up to date on an ongoing basis, it is preferred that the USPTO provide a Web API or service that we can use to retrieve a list of applications that have been added to or updated in Public PAIR within a given time period. An alternative to an API or service would be a file that would be generated by the USPTO and delivered to us on a daily basis. [redacted]

The crawling process will be designed to capture, create and host [redacted] new patent applications per day (however, note the contractor's need for understanding USPTO capacities, as referenced in Section 2.1.2.2 above; for purposes of this proposal we assume that agency capacity constraints will not be an issue) . [redacted]

***2.1.2.2.2*** [redacted]

[redacted]

- [redacted]

[redacted]

**2.1.2.3** [redacted]

[redacted]

[redacted]

**2.1.2.4 Data cleanup**

**(Reference RFP C.3.2.1, C.3.13)**

The data cleanup requirement involves the ability to retrieve missing or replacement bulk data files from the USPTO when available. [redacted]

**2.1.3 Repository**

**(Reference RFP C.3.1, C.3.8)**

The free public access service would be hosted [redacted]

Please refer to the following diagram as the attributes of the data environment are discussed:

**Data Environment**

### 2.1.3.1 Data security

**(Reference RFP C.3.16, C.3.18, C.4.1)**

Reed Tech will build controls around the USPTO data to ensure the integrity and availability of the data and to prevent unauthorized access to the system.

- [redacted]

### 2.1.3.2 Location
### (Reference RFP C.3.18)

[redacted]

Reed Tech will notify the USPTO within 24 hours of any issues that render the data dissemination service unavailable for public access. Additionally, Reed Tech will detail our remediation or mitigation plan to return the service to operational status.

### 2.1.3.3 Redundancy

Reed Tech's proposed solution will ensure adequate responsiveness of the application to the public and continuous availability by implementing [redacted]

[redacted]

### 2.1.4 Public interface and retrieval

#### 2.1.4.1 Website portal

**(Reference RFP C.3.5, C.3.7)**

Reed Tech will design, set up and maintain a website portal to provide public access to download patent and trademark bulk data files plus the ability to search for and download individual patent application PAIR records from the repository. The website will operate 24/7 and 1) will be accessed by its own domain URL, 2) [redacted]

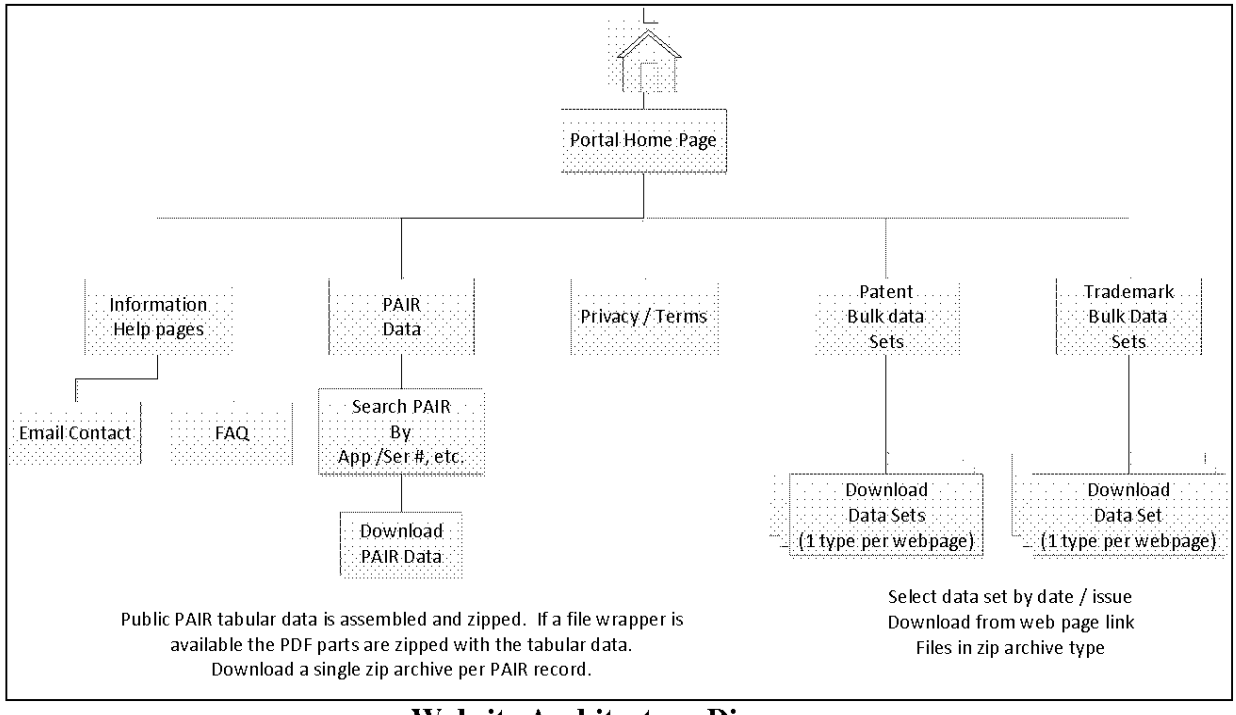
Reed Tech will ensure that at least [redacted] internet bandwidth is available for public download of bulk patent data. With this configuration, as much as [redacted] of data can be downloaded each day.

#### 2.1.4.2 Website navigation

The following primary pages will be included in the site:

1. Home: Portal landing page with links to all products, information messages and update status.
2. Trademark Bulk Data: Main page with links to sub-pages for downloading individual Trademark data files as listed on RFP Attachment 1, Current Data Sets. Date range, document range and content lists will be provided to assist the public in identifying files for download.
3. Patent Bulk Data: Main page with links to sub-pages for downloading individual patent data files as listed on RFP Attachment 1, Current Data Sets. Date range, document range and content lists will be provided to assist the public in identifying files for download.
4. Public PAIR application search and download: Page with instruction and search entry to allow user to enter and download zipped PAIR application data. Search entry will allow the user to enter an application serial number, USPTO control number, publication number or patent number to search the repository and retrieve a zipped file link that can be downloaded.
5. Help/FAQ: Page with documentation for all data types, data download instructions, common problems/solutions and a feedback/report problem form.
6. Privacy and Terms: Pages with appropriate policies and terms statements.

These points can be seen in the following diagram:




**Website Architecture Diagram**

### 2.1.4.3 Availability

**(Reference RFP C.3.10)**

As discussed above, the full technology stack will contain redundant configurations to help ensure continuous service to the public. We are designing the system to maintain [redacted]

### 2.1.4.4 Help desk support

Reed Tech currently maintains an internal help desk that supports a variety of external and in-house customers. Reed Tech will offer support to the public via email. Service requests will be logged and tracked through our ticketing system. It is anticipated that all contacts will be responded to via email within four hours of receipt during normal working hours (Monday-Friday, 8AM-8PM ET). Requests received off-hours will be responded to during the next business day.

### 2.1.5 Reporting and analytics

**(Reference RFP C.3.2.1, C.3.17)**

In addition to the standard project-related reporting defined in RFP Section F.3 Contract Deliverables, Reed Tech will also [redacted]

[redacted]

Additionally, Reed Tech can offer the USPTO [redacted]

## 2.2 Schedule milestones

Presented here is an initial draft of a potential schedule with high level milestones. This plan is subject to revision pending discussions with the USPTO following contract award. It is also envisioned that access to USPTO systems will be provided during system development to enable testing.

Because of dependencies associated with USPTO systems, discussions will be necessary to validate assumptions in the draft timeline below. Information about Reed Tech skill sets and resources necessary to develop and implement the environment are provided in the Experience and Resources section below.

[redacted]

## 2.3 Program management approach

Reed Tech has extensive experience in managing a large contract for the USPTO.

Reed Tech's program management approach is based upon the industry-recognized project management processes of the Project Management Institute (PMI). Reed Tech's approach is organized by the PMI's five Process Areas (initiating, planning, executing, controlling, and closing). In support of its program management practices, Reed Tech has several PMP® and CSM® credential holders on staff.

Aligned with the PMI's philosophy, RTIS distinguishes between the Project Management Life Cycle (the processes that specify how projects are managed) and the Product Development Life Cycle (the processes that specify and create actual products, e.g., software development applications and tools). For software development projects, a determination of the software development process is made based upon factors such as the size and complexity of the project and the product's compliance requirements.

Reed Tech's Project Life Cycle encompasses the planning and control of resources within the five process areas described below:

1. Initiating – Upon contract win, the project is initiated, approved and begun through a formalized request process.
2. Planning – A program manager is assigned with responsibilities for:

    - Generation of a project management plan that describes project implementation. This includes: how and when activities will be undertaken as well as the procedures that will be followed to ensure compliance with contract requirements; escalation to management of schedule problems; assigning roles and responsibilities and specifying reporting requirements per the contract.
    - Working with Technical Manager, Designated Developers and Subject Matter Experts to estimate Level of Effort (LOE) and resource allocation for the project.
    - Generating a project schedule (or schedules) based on inputs from stakeholders and contract requirements.
    - Overseeing and coordinating changes affecting the scope of the program and ensuring that changes are accurately defined and estimated.
    - The overall communications regarding the program. They are the primary single authoritative source of information and decision making to ensure a shared understanding of all program parameters.

3. Execution – The Technical approach described in Section 2 of this RFP is executed with oversight from the program manager.
4. Monitoring and Controlling – The program manager monitors in-process and completed work results to ensure adherence to the project plan. Plan variances and their respective corrective actions are escalated to appropriate management personnel and, if required, to the customer.
5. Closing – When all scheduled project tasks are completed and accepted, the program manager closes out the project and creates and issues a final project report.

# 3 Experience and resources

## 3.1 References and associated experience

### 3.1.1 Reed Tech experience with USPTO data and reference

Continuously since 1970 Reed Technology and Information Services Inc. ("Reed Tech") has provided data capture, processing and publication services to the USPTO. Reed Tech has prepared much of the patent bulk data and even some of the trademark data included in this solicitation. Some of the areas Reed Tech has participated in are:

- 1970 to present: Grant full text and bibliographic data capture and publication, including preparation of the files used by the Office for data dissemination;
- 2001 to present: Application full text data capture and publication, incorporating application bibliographic data with the full text data into the data files prepared for dissemination;
- Regular participation in internal and international meetings to create and modify the specifications for the XML data files;
- 2005 to present: Image capture and Doccode assignment of application-related papers received by the office for loading into IFW;
- 2007 to present: Review of EFS documents submitted by applicants to ensure accuracy of Doccodes assigned; and
- 1980 to 1984: capture of trademark applications published for opposition, marks registered, renewals, cancellations, etc.

In addition, Reed Tech:

- [redacted]

Reed Tech developed its patent data capture and document management system internally using a wide variety of COTS tools integrated with custom in-house developed software. We have significant expertise in the software development process, document management, OCR, image processing, workflow, data validation, and quality management.

[redacted]

[redacted]

### 3.1.2 PDR3D experience and reference

Reed Technology designed, created and maintains the PDR3D online drug label database system for PDR Network. PDR3D is a digital reference and search tool that contains over 75 thousand drug labels including branded and generic prescription drugs, over the counter drugs, homeopathic drugs and biologics. The databases are updated daily with the most current drug label information. The Web based search tool has the following primary features:

- Daily updates are loaded from the FDA/NLM datasets.
- Robust search capabilities: Users can easily search by drug name, NDC code or build advanced searches using 11 fields with over 90 different search parameters. Also search labels based on pharmacologic class.
- Label Tracking: LabelTrak allows you to easily stay up-to-date with frequently changing human drug labels. Set up alerts for targeted products and the LabelTrak system will automatically monitor for updates to those products. When updates are made to a label, LabelTrak will automatically notify you of those updates quickly and conveniently via email.
- Label Compare: Easily compare labels and label parts even down to small SPL/XML changes.
- Exporting: Search results can be exported to Excel or PDF.
- Easy access to all daily label updates.

The PDR3D system web servers and database are housed in the Reed Tech datacenter. [redacted]

[redacted]

### 3.1.3 TotalPatent™ experience and reference

TotalPatent is a state-of-the-art, easy-to-use patent search, retrieval and workflow solutions tool, created by Reed Tech's parent, LexisNexis. LexisNexis resources are available to Reed Tech as needed. It is used by law firms, corporations, government and academic institutions. The product offers the following features:

- Access to:
  - >40 terabytes of patent data
  - >330 million images
  - >90 million patent documents
  - >65 million PDFs, compressed to 1/3 of their original size
  - >35 million English full text translations (machine translations of foreign patent documents)
  - 100 authorities, of which 31 are full text searchable in English, including publications from the IP5 (US, EP, CN, JP and KR) offices
- US data is available on the day of publication and includes over 10 million high-quality front-page –clipped images" for display and download.
- Utilizes powerful, fully transparent and user-controllable latent semantic analysis (LSA) in addition to Boolean search capability (finds patents a regular Boolean search may not).
- Over 56 million main patent families, with the ability to 1) easily remove patent family duplicates prior to search, 2) sort by patent family after search and 3) produce patent family reports
- Nearly 200 indexed and searchable patent fields
- Ability to retrieve English translations of foreign language patents/applications; all links are in English, even in non- English documents.
- Legal status information for 55 authorities, including important information such as reassignment, reissue or lapse information
- Easy to use web-based tool with 24/7/365 USA-based support

[redacted]

[redacted]

[redacted]

[redacted]

### 3.2 General resources

With nearly [redacted] employees and over [redacted] person-years of patent conversion and tagging experience on its staff, Reed Tech has the proven technical and operational knowledge, experience and capacity. Reed Tech has developed systems to ingest, convert, and tag over 4.5 million patents for the USPTO since 1970, handling all document and content types. The Reed Tech executive team has a combined 100+ person years of experience in creating and delivering solutions for patent data management.

Reed Tech's technology team, consisting of approximately [redacted] ndividuals, is staffed by experts in every aspect of patent content management. This team built and currently supports an environment for Reed Tech's PaDaCap contract with the USPTO including automated workflow and management of more than [redacted] of data.

[redacted]

Reed Tech's IT engineering team possesses expert knowledge in the areas of redundant networking and system design, data management, storage and data replication technologies, virtualization, database management, identity and access management. The Reed Tech infrastructure currently consists of:

- [redacted]

### 3.3 Specific skills and capabilities

Reed Tech possesses skill sets across a wide spectrum. Our Technical Team, a group of approximately [redacted] with well over [redacted] person-years of experience, includes software architects, developers, QA test specialists, validation specialists, database administrators, system engineers, security engineers and system operations specialists. Our Program Management and Project Management team, a group of [redacted] with over [redacted]-person years of experience, includes

program/project managers, requirements specialists and business analysts. Some of the particular capabilities held by Reed Tech staff and applying to this contract are seen in the following position descriptions.

**3.3.1** [redacted]

[redacted]

**3.3.2** [redacted]

[redacted]

[redacted]

### 3.3.3 [redacted]

[redacted]

[redacted]

### 3.3.4 [redacted]

[redacted]

- [redacted]

### 3.3.5 [redacted]

[redacted]

[redacted]

### 3.3.6 [redacted]

[redacted]

[redacted]

### 3.3.7 [redacted]

[redacted]

### 3.3.8 [redacted]

[redacted]

### 3.3.9 [redacted]

[redacted]

# 4 Other RFP requirements (Reference RFP C.3.19, C.3.20)

## 4.1 Return of data

Reed Tech will return all data associated with this contract to the USPTO on hard disk or other agreed upon media within 90 days of contract termination.

## 4.2 USPTO-17 Contractor FOIA

Reed Tech will supply a version of contract to the USPTO within 14 days of contract award that makes changes we believe are necessary or are required by law.

# 5 Summary

This proposal describes Reed Tech's extensive capacity and technical capabilities to perform the work described in the RFP. Moreover, we have very deep domain knowledge of the US patent data that is the subject of the RFP, and our 42+ year history of patent data capture, document management and service to the USPTO underscores our deep commitment to helping the agency achieve its objectives.

We believe we offer the ability to implement the program requested in this solicitation in a manner that will minimize USPTO involvement and investment of resources in it. We also see it as an opportunity to serve the intellectual property community, and we expect to solicit ongoing user feedback in order to determine the best ways to improve public access to this extremely important content. We have the resources, domain expertise, and the commitment to add value to the USPTO's data dissemination processes in a manner that will serve both the USPTO and the public. We stand ready to support the USPTO in this important task.

# Volume II

# Section B

United States Patent and Trademark Office (USPTO) Request for Proposal (RFP) Number:
DOC50PAPT1300005
Public Dissemination of Data

**Section B – Supplies or Services**

**B.1 Contract Type**

The USPTO intends to award one (1) no-cost contract. Evaluation will be on the Base and all Option Periods.

**B.2 Period of Performance**

This contract award will consist of a base period of one (1) year and four (4) one-year option periods. The period of performance for the contract will begin on the date of award.

**B.3 Schedule of Values**

**Base Period**

| Contract Line Item Number (CLIN) | Description | Units | |
|---|---|---|---|
| 0001 | Public Dissemination of Patent and Trademark Bulk Data | 1 Lot | No cost to USPTO |

**Option Period 1**

| Contract Line Item Number (CLIN) | Description | Units | |
|---|---|---|---|
| 1001 | Public Dissemination of Patent and Trademark Bulk Data | 1 Lot | No cost to USPTO |

**Option Period 2**

| Contract Line Item Number (CLIN) | Description | Units | |
|---|---|---|---|
| 2001 | Public Dissemination of Patent and Trademark Bulk Data | 1 Lot | No cost to USPTO |

United States Patent and Trademark Office (USPTO) Request for Proposal (RFP) Number:
DOC50PAPT1300005
Public Dissemination of Data

**Option Period 3**

| Contract Line Item Number (CLIN) | Description | Units | |
|---|---|---|---|
| 3001 | Public Dissemination of Patent and Trademark Bulk Data | 1 Lot | No cost to USPTO |

**Option Period 4**

| Contract Line Item Number (CLIN) | Description | Units | |
|---|---|---|---|
| 4001 | Public Dissemination of Patent and Trademark Bulk Data | 1 Lot | No cost to USPTO |

**B.4 CLINs 0001- 4001**

The CLINs and their associated Option CLINs shall include all necessary labor, materials, equipment and other resources necessary to provide the specific deliverables identified in Section F.3. The Government will not reimburse, or provide for any expenses, including travel expenses.

# Amendment I

## Public Dissemination of Data RFP – Questions and Answers

| No. | Question | Response |
|---|---|---|
| 1 | Page 6 of the RFP states a requirement that, —under the specific direction of the USPTO" the contractor shall —provide data cleanup where necessary, including but not limited to the identification and correction of data, elimination of duplicates and the correction of inconsistent data…."<br><br>*(Questions in reference to this statement are #1.1-1.3)* | NA – |
| 1.1 | 1. Can the USPTO offer some indication of how extensive this data cleanup requirement will be (i.e., what volume of data is expected to require cleanup, and with what frequency)? | The USPTO averages a missing or corrupt Trademark Daily File twice a month. The correction involves replacing the file in its entirety. A similar problem occurs once or twice a year with the Patent or Patent Application weekly files. |
| 1.2 | 2. Will the USPTO identify specific errors for the contractor to clean up, or will error identification be the contractor's responsibility? | The USPTO will identify the specific errors. |
| 1.3 | 3. Can the USPTO provide specific examples of data cleanup that will fall under this requirement?" | For example, the Trademark 24 Hour Box Supplemental file for a specific date, hrs130101a.zip, was not available for download (missing). Resolution: the file is provided on the USPTO site and subsequently the contractor's site. |

| No. | Question | Response |
|---|---|---|
| 5 | Approximately how much storage space is required to house the current data covered by the RFP and what is the projected growth rate per year? | Currently, including Public PAIR data, the total collection is estimated at 100 TB, with an anticipated annual growth of 10 TB. Please see the USPTO RFP_Attachment_1_Current_Data_Sets_20120425.doc that gives specific collection file sizes and annual file sizes. |
| 6 | Does the USPTO have any requirements regarding the number of hours that the Contract Administrator and CA's Technical Representative are available or working on this project? | NA - USPTO understands this question to be related to USPTO support; USPTO will provide the necessary resources and support to this project. |
| 7 | How will the availability of the service be advertised or made known to the public, will this responsibility fall to the USPTO or be carried by the contractor? | The USPTO will be responsible for advertising the service and will do so via web and email announcements, Data.Gov, and links from the USPTO website. |
| 8 | Regarding the "XML Schemas, Style sheets, data dictionary, containers, components and all artifacts necessary to define the data for the purposes of dissemination" will there be cooperation with the contractor, and will suggestions be received from the contractor, regarding the modification to or enhancement of these definitions for the purposes of system optimization? | The USPTO would entertain suggestions from the contractor regarding modification to or enhancement of the definitions of the data. However, several of the USPTO (patent and patent application) XML formats are controlled by international committee (European Patent Office, Japan Patent Office, etc.) and are not easily modified. |

| No. | Question | Response |
|---|---|---|
| 9 | Since any enhanced content will be considered "official USPTO Licensed Content" will the contractor be required to pay any portion of monies received by the sale of access to the enhanced content to the USPTO or is that revenue kept entirely by the contractor? | Enhanced content is not considered official USPTO Licensed Content. See C.1.3.7: —Asa result of any award, any repackaging, value added content, or enhancements to the Licensed Content (see definition below) created by the Contractor shall not be considered to be official USPTO Licensed Content and the USPTO is neither responsible nor liable for the accuracy of such repackaged, value added, or enhanced products." Any revenue received by the contractor for use of enhanced content is entirely the contractor's. |

## Representations and Certifications

Current Representations are in ORCA and SAM.