



UNITED STATES PATENT AND TRADEMARK OFFICE



Public Meeting On Data Dissemination Request for Information

Office of the Chief Information Officer
September 24, 2009



Agenda



- Introduction
- Overview
 - ▶ Data Dissemination Objectives
 - ▶ Current Limitations
- Key Objectives of RFI
- Questions



Federal Data Dissemination Objectives

- **Transparent**
 - ▶ Promote accountability
 - ▶ Provide information for citizens on what their government is doing
- **Participatory**
 - ▶ Agencies encouraged to provide citizens opportunities to participate in policy making
 - ▶ Agencies encouraged to solicit ideas from citizens how to improve those opportunities
- **Collaborative**
 - ▶ The public builds innovative tools to enable collaboration across and at all levels of government and private industry



Current USPTO Data Delivery Options



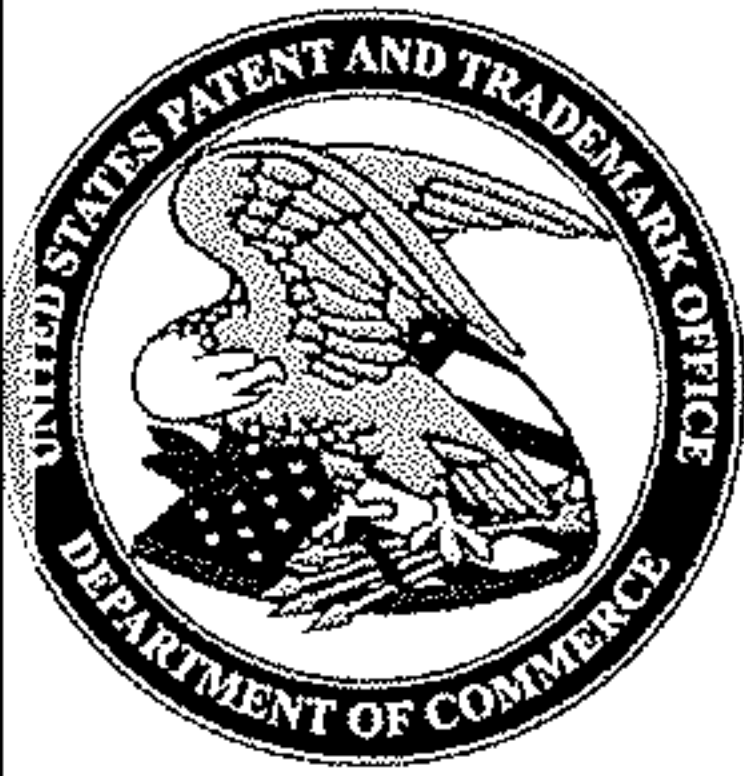
- USPTO has over 150 distinct data products comprising
 - ▶ Bulk data sets
 - ▶ Interactive optical disc products
 - ▶ Statistical reports
 - ▶ Details of these data sets are in Attachment 1 – Current Data Sets

- Free access to raw data sets through Data.gov
 - ▶ Patent Bibliographic Data for Grants and Applications
 - ▶ Trademark Application Text and Images
 - ▶ Trademark Trial and Appeal Board data
 - ▶ Trademark Assignment data



Current USPTO Data Delivery Options, continued

- Online search systems designed for Single Queries, including:
 - ▶ AppFT/AIW and PatFT/PIW – Patent Grants and Applications
 - ▶ Public/Private PAIR – Patent file wrapper and status information
 - ▶ PSIPS – Patent Sequencing Data
 - ▶ AOTW – Patent and Trademark Assignments
 - ▶ TTAB - Trademark Trial and Appeal Board
 - ▶ TARR - Trademark Applications and Registrations
 - ▶ TDR - Trademark File Wrapper
 - ▶ TESS - Trademark Electronic Search System (TESS)
- These systems were not designed for large numbers of hits, bulk queries and deliveries, and are directly linked to examiner's systems.



Key Requests from IP Community



- More Data – All of it!
- Unrestricted Access – as much as I can consume
- Flexible Delivery Options – Bulk
- Consistent Formatting – Useful data format
- Public Access to Data at no cost
- Fairly Distributed



Data Dissemination Limitations

There is a plan to fix this over the next six years – USPTO wants to accelerate Data Dissemination in the Public interest

- Hardware & network – Industry standard average lifespan is 5-7 years
 - ▶ Almost all the network switches were purchased in 2000
 - ▶ 19% of our servers are 7 years or more
 - ▶ Beyond the 7 year lifespan – no support for problems, no replacement parts, & growing failures
 - ▶ Incredible growth in the number of Examiners has maxed out our systems
- Software
 - ▶ Legacy Software running on aging and outdated technologies
 - ▶ Unmanaged growth and change
 - ▶ Poor documentation of configurations and interconnectivity
 - ▶ Deficient coding standards
- Data structure – integrated public and private data, making preliminary extract necessary
- Data volume – approx. 10,000 patent grants and published applications each week PLUS metadata



Data Dissemination Limitations



Data not yet available in bulk

- PALM Meta Data of Published Applications or Patents
 - ▶ Transaction History
 - ▶ Patent Term Adjustment
 - ▶ Foreign Priority
 - ▶ Attorney Address
 - ▶ Foreign Priority
 - ▶ Continuity Data
 - ▶ Publication Info
- IFW Meta Data associated with the Documents
- IFW Image Data
- All Fee Information for Publish Applications
- Trademark retrospective file of cropped images/logos



Overview

High Level Parts

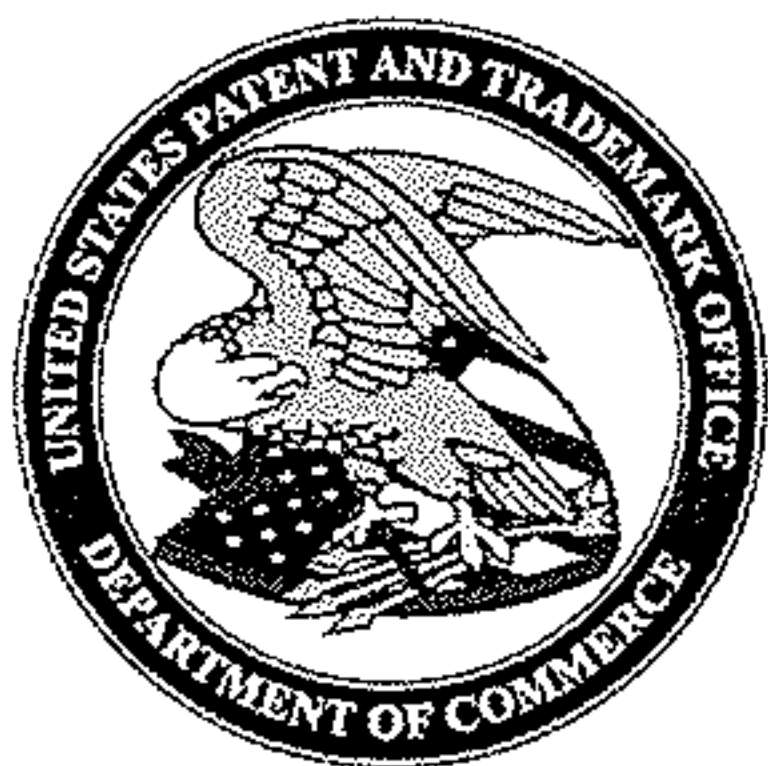
- Data extraction (at the USPTO)
 - ▶ Data scattered through many UPSTO systems and data stores
 - ▶ Private and Public data mixed
 - ▶ Systems tied to examiner production (must separate)
 - ▶ Have to honor Federal IT security standards
- Data Processing/Packaging (at the USPTO)
 - ▶ Need a smaller foot print of data than there is today
 - ▶ Image and text in the same package
 - ▶ Various formats today; USPTO needs to embrace open standard formats (e.g., international standards, XML, PDF, & OpenDoc)
- Data Hosting (Anywhere)
 - ▶ Outside of USPTO
 - ▶ Equal & Fair (Public data distributed quickly to all before enhanced data)
 - ▶ Bulk data ('Checksummed' as-is) for free



Key Objectives of RFI



- Market Research
- No Cost Contract to the Federal Government (USPTO)
- Fund development of new infrastructure
- Work with USPTO OCIO – OCIO will own the system, including all data rights and associated code
- Fund ongoing operations; access to new data
- Data format recommendations
- Redistribute the data as-is from USPTO to the public for free
- Add value to the data and sell it, provide access tools, etc.



RFI Responses

- Written responses only
- 20 page limit
- Electronic submission only
- Send email with responses to this E-mail box by 2:00 pm, EST on October 15, 2009:
publicdatadissemination@uspto.gov
- RFI modification to be posted on FBO with Q&A



Question and Answer Document



All of the questions provided to USPTO prior to today's meeting are provided in the handout along with the agency's responses.



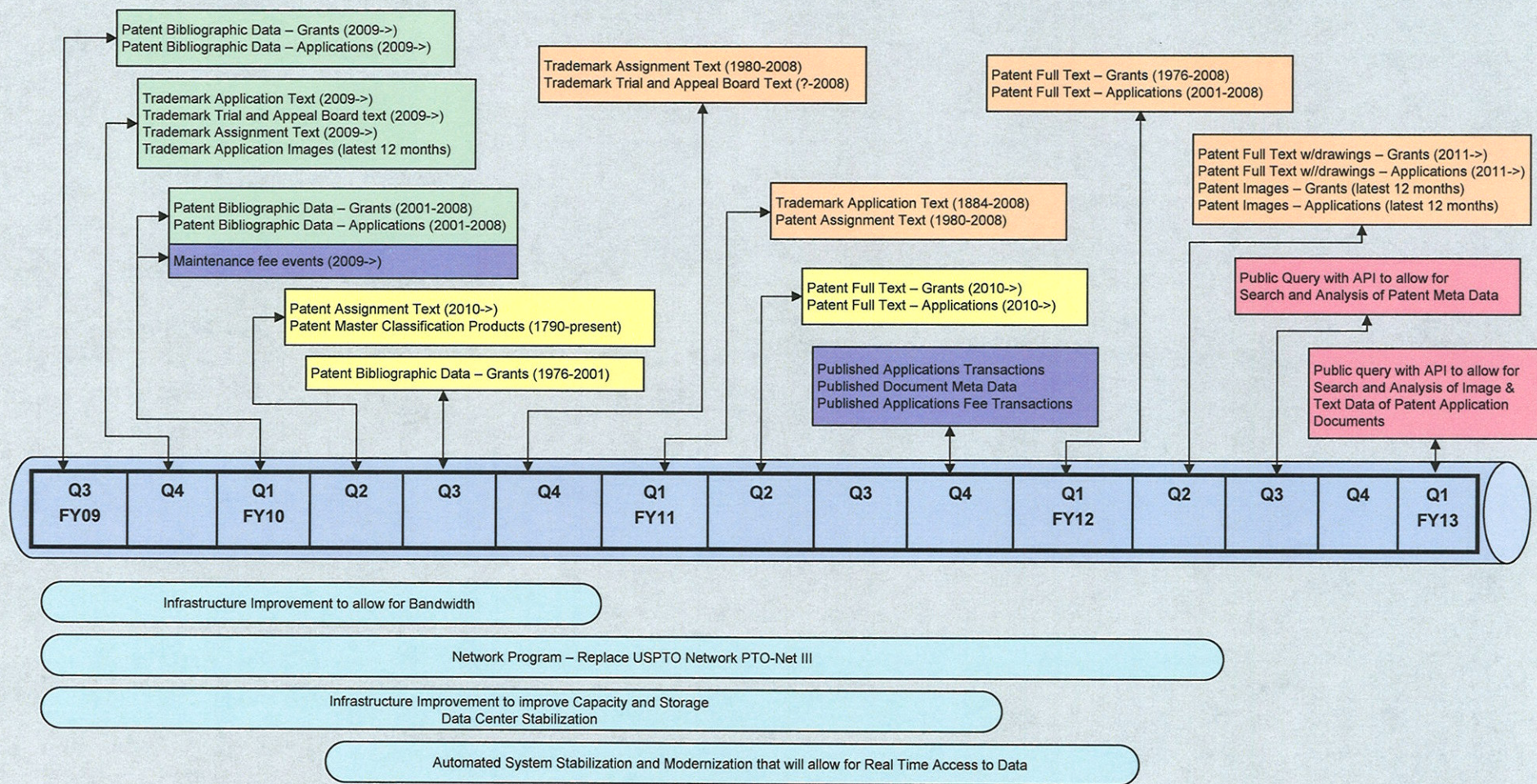
Q&A

Questions?

Data Set	Description	Issue Frequency	Distribution Format	Approx. Issue Size *	Collection Sizes
Patent Bibliographic Data – Grants (2009->)	Contains the bibliographic text (i.e., front page) of each patent grant issued weekly (Tuesdays) for 2009 and later.	Weekly	XML	5 MB	260 MB (per year)
Patent Bibliographic Data – Grants (2001-2008)	Contains the bibliographic text (i.e., front page) of each patent grant issued weekly (Tuesdays) for 2001-2008 in various formats.	Single	SGML, XML	N/A	2 GB
Patent Bibliographic Data – Grants (1976-2008)	Contains the bibliographic text (i.e., front page) of each patent grant issued weekly (Tuesdays) for 1976 through 2008.	Single	ASCII, SGML, XML	N/A	5.0 GB (1976-2008)
Patent Bibliographic Data – Applications (2009->)	Contains the bibliographic text (i.e., front page) of each patent published application issued weekly (Thursdays) for 2009 and later.	Weekly	XML	2.7 MB	140 MB (per year)
Patent Bibliographic Data – Applications (2001-2008)	Contains the bibliographic text (i.e., front page) of each patent published application published weekly (Thursdays) for 2001-2008 in various formats.	Single	XML	N/A	1.1 GB
Trademark Application Text (2009->)	Contains pending and registered trademark text data.	Daily	XML	17 MB	6.2 GB (per year)
Trademark Application Text (1884-2008)	Contains pending and registered trademark text data.	Single	XML	N/A	9 GB (1884-2008)
Trademark Trial and Appeal Board (TTAB) text (2009->)	Contains text data for proceedings of the Trademark Trial and Appeal Board (TTAB).	Daily	XML	240 KB	88 MB (per year)
Trademark Trial and Appeal Board (TTAB) text (?-2008)	Contains text data for proceedings of the Trademark Trial and Appeal Board (TTAB).	Single	XML	N/A	164 MB (?-2008)
Trademark Assignment Text (2009->)	Contains trademark assignment text derived from assignment recordations made at the USPTO for 2009 and later.	Daily	XML	46KB	17 MB (per year)
Trademark Assignment Text (1980-2008)	Contains trademark assignment text derived from assignment recordations made at the USPTO for 1980-2008.	Single	XML	N/A	126 MB (1980-2008)
Trademark Application Images (latest 12 months)	Contains images of Trademark applications in JPG (black and white, grey-scale, or color) and TIFF (black and white) format processed daily.	Daily	XML, TIFF, JPEG	215 MB	56 GB
Patent Assignment Text (2009->)	Contains patent assignment text derived from assignment recordations made at the USPTO for 2009 and later.	Daily	XML	2.3 MB	825 MB (per year)
Patent Assignment Text (1980->2008)	Contains patent assignment text derived from assignment recordations made at the USPTO for 1980-2008.	Single	XML	N/A	1.7 GB (1980-2008)
Maintenance Fee Events (2009->)	Contains patent maintenance fee event information. Created Mondays and available Tuesdays.	Weekly	ASCII	62 MB	3.2 GB (per year)
Patent Master Classification Products (1790-present)	Current U.S. classification information for all patent grants issued by the U.S. Patent and Trademark Office from 1790 to present.	Bimonthly	ASCII	261 MB	1.5 GB (1790-present)
Patent Full Text – Grants (2009->)	Contains the full text of each patent grant issued weekly (Tuesdays) for 2009 and later; excludes images/drawings.	Weekly	XML	54 MB	2.8 GB
Patent Full Text – Grants (1976 To 2008)	Contains the full text of each patent grant issued weekly (Tuesdays) for 2001 to 2008; excludes images/drawings.	Single	ASCII, SGML, XML	N/A	92 GB (1976-2008)
Patent Full Text, with drawings – Grants (2009->)	Contains the full text, drawings, and complex work units (tables, mathematical expressions, genetic sequence data, and chemical structures) of each patent grant issued weekly (Tuesdays) for 2009 and later.	Weekly	XML	836 MB	43.4 GB (per year)
Patent Full Text – Applications (2009->)	Contains the full text of each patent application (non-provisional utility and plant) published weekly (Thursdays) for 2009 and later; excludes images/drawings.	Weekly	XML	85 MB	4.4 GB
Patent Full Text – Applications (2001-2008)	Contains the full text of each patent application (non-provisional utility and plant) published weekly (Thursdays) for 2001 to 2008; excludes images/drawings.	Single	XML	N/A	35 GB (2001-2008)
Patent Full Text, with drawings – Applications (2009->)	Contains the full text, drawings, and complex work units (tables, mathematical expressions, genetic sequence data, and chemical structures) of each patent application (non-provisional utility and plant) published weekly (Thursdays) for 2009 and later.	Weekly	XML	1.5 GB	78 GB (per year)
Patent Images – Grants (1790-present)	Contains the images of each patent grant issued weekly (Tuesdays).	Weekly	Single-page TIFF images	4 GB	208 GB
Patent Images – Applications (2001-present)	Contains the images of each patent application (non-provisional utility and plant) published weekly (Thursdays).	Weekly	Single-page TIFF images	6 GB	312 GB

	Description	Update Frequency	Source Format	Approx. Update Size	Collection Sizes
Patent Application Meta Data	Contains patent application meta data corresponding to the tabs in PAIR, including "Application Data", "Transaction History", "Patent Term Adjustments", "Patent Term Extension History", "Foreign Priority", "Address & Attorney/Agent", "Continuity Data" and "Published Documents".	Daily	ASCII	11 MB	67 GB
Patent Application Fee Transactions	Contains patent application fee transactions	Daily	ASCII	5 MB	30 GB
Patent Application Document Meta Data	Contains patent application document meta data corresponding to the "Image File Wrapper" tab in PAIR, including "Document Description", "Mail Room Date" and "Page Count".	Daily	ASCII	80 KB	96 MB
Patent Application Documents	Contains patent application documents corresponding to the documents available in the "Image File Wrapper" tab in PAIR, including Specification, Claims, Amendments, etc.	Daily	TIFF images	42 GB	50 TB

Target Timeline for USPTO Participation in Data.gov



Assumptions:

Combined funding and progress on USPTO/CIO Roadmap and Strategic IT Plan
Adequate disk space
Sufficient network and internet bandwidth
Legal rights to charge only for media delivery

Product Key:

Free	Currently Online	New Online Product
	Currently Media-based	New Online Tool

Public Data Dissemination RFI Questions and Answers
(Represents questions received through 2:00 pm, September 22, 2009)

No.	Question	Answer
1	USPTO delivers various data in limited quantities to the public today. Will USPTO continue to perform this function once data is delivered to a contracted partner? Will USPTO dissemination also be at no charge to the public?	The goal is to enter into a partnership where the awardee(s) remove the stress of data delivery from the USPTO infrastructure by delivering the data to the public. However, USPTO will continue to sell packaged data as long as there is a demand for that service.
2	Will third party partners (awardees) be required to enhance the data they receive from USPTO before providing it to the public for free? If yes, what are the requirements?	No. Awardees must pass the data that USPTO provides to them on to the public, as is, at no charge. They may also enhance the data or delivery mechanism and charge for the enhancements.
3	If the data that USPTO provides to awardees is passed on to the public unaltered, will USPTO continue to be accountable for the accuracy of the data?	Yes. Legal responsibility for the accuracy of the unaltered data remains with USPTO.
4	This RFI seems focused on PAIR data. Does USPTO plan to disseminate other, existing data sets, such as patent full-text, via this same mechanism? If so, will they also be distributed to the public at no charge?	Yes, USPTO intends to use this mechanism to disseminate all public data, including existing fee-based data sets via this mechanism, at no charge to the public. However, USPTO will continue to sell packaged data as the agency does today for as long as there is a demand for the packaged data.
5	Does USPTO plan to scan or convert paper files into electronic form? Does USPTO expect awardees to do that?	USPTO has no plans to convert existing paper-based files to electronic form at this time. USPTO encourages awardees to include plans to perform this function if they believe that doing so provides a benefit to the awardee and/or the public.
6	What kind of technical support would awardees be expected to provide?	Data expertise on the data that is provided by USPTO would be provided directly to customers by the USPTO. Data that has "added value", as well as technical assistance with delivery mechanisms would be part of the awardees' responsibility.
7	What is the historical and projected percent growth per year of the data sets?	Attachment 1, Current Data Sets provides the size and growth data for each current and prospective data set. USPTO encourages offerors to propose, in their responses to the RFI, methods for reformatting the data to reduce storage and bandwidth requirements for subsequent distribution.

No.	Question	Answer
8	Does the USPTO expect to add additional data sets in the future? How would these be handled in the contract?	USPTO intends to make all of the existing data sets available under this arrangement and we do not currently anticipate the introduction of new data sets. Handling of any new data would have to be addressed through subsequent contracting actions.
9	Is the vendor expected to process, validate, correct, or improve, any of the data provided by the government for free dissemination?	<p>No. The data passed on to the public does not need to be modified in any way but must be provided to the public at no charge.</p> <p>However, USPTO encourages offerors to add value to the data and distribute the enhanced data for profit.</p>
10	Has/will the USPTO considered opening this opportunity under an 8(a) set-aside?	The USPTO will consider setting this opportunity aside if it is determined that at least 2 -8(a) vendors are capable. At this time an acquisition strategy has not been made.
11	The RFI indicates that responses will only be accepted in Microsoft Word. Is there any reason that responses cannot be accepted in other standard industry formats such as OpenOffice or PDF? Has Microsoft furnished any monetary or in-kind compensation as part of the requirements for use of Microsoft Word, Microsoft Visio, and Microsoft Excel for submissions?	<p>The Microsoft Office suite of products is USPTO's standard tool set. Accepting electronic submissions in other formats may inhibit the agency's ability to read the files. Responses may be provided in PDF format as well.</p> <p>Microsoft has not provided compensation of any kind relative to this RFI.</p>
12	It was unclear if all bids will be made public after the redaction of any unmarked proprietary information. Knowing what bids have been submitted will allow the public a greater ability to understand the tradeoffs made by the U.S. Patent and Trademark Office in evaluating the submissions and for any subsequent decisions to issue an RFP.	Information submitted as part of the RFI is for market research and planning purposes of the USPTO only. The office does not intend to make the submissions public any responses to the RFI that are marked proprietary or confidential
13	Given the fundamental and long-lasting repercussions of this initiative, has any thought been made to holding a West Coast or Midwest open meeting in addition to the one being held inside the Washington Beltway? Likewise, will audio and video from the meeting be made available to the public on the Internet?	USPTO will consider alternatives for this request.
14	As this initiative may fundamentally alter your distribution strategies, will the Under Secretary and other members of the U.S. PTO senior management team also be participating in this process?	USPTO senior management will participate in the process.

No.	Question	Answer
15	<p>Can USPTO provide more details of the 2 petabytes of data:</p> <p>a. Details of the data formats and volumes for each data set that is available in electronic format?</p> <p>b. Details of the data sets that are not currently in electronic format</p>	<p>Yes. Attachment 1 to the RFI contains a broader description of the data being considered within the scope of this effort.</p>
16	<p>Can USPTO provide an overview of its current data dissemination platforms and operations?</p>	<p>Current bulk data products are distributed via download, magnetic tape, or optical disc, depending on factors such as size and frequency. Some products are provided as an annual subscription. While some data sets are readily produced as part of current data processing operations, others would have to be periodically extracted from internal databases.</p>
17	<p>What is the current approach to timing for making bulk data available to the public and to commercial vendors?</p> <p>What are the goals and objectives of the PTO RFI as it relates to these data sets? (status quo, versus enabling data mining, versus additional system enhancements)</p>	<p>Attachment 2 to the RFI represents the current plan for making USPTO data available to the public through data.gov.</p> <p>The goal of this effort is to enhance the accessibility and usability of USPTO data in an accelerated manner. There are two objectives: 1) to offer no-charge access to current bulk data sets; 2) to offer no-charge access to data that is not currently available in bulk (e.g., PAIR)</p>
18	<p>What are USPTO's initial expectations for Service Level Agreements, Quality Control, and Accessibility?</p> <p>a. Requirements for uptime / peak access</p> <p>b. Scheduled maintenance</p>	<p>Certain data sets must continue to be provided at weekly intervals (Tuesdays for Patent Grants, Thursdays for Patent Applications). Other data should be distributed at least as frequently as currently done: distributed daily (Trademark Applications, and Trademark Trial and Appeal Board data, as well as Trademark and Patent Assignment data); or bimonthly for classification data. Attachment 1 to the RFI contains the issue frequency of current data sets.</p> <p>Service level agreements and requirements to include quality and accessibility have not yet been finalized. These requirements will be incorporated into any subsequent procurement actions.</p>
19	<p>Should data exchange with other Patent Offices, including the delivery of data prior to the official publication date, be covered by any proposed Data Dissemination solution?</p>	<p>No data is disseminated prior to the official publication date. The USPTO exchanges data with a limited number of other Patent Offices (< 10), and would consider using a 3rd-party solution for that delivery.</p>

No.	Question	Answer
20	Would all components of a data dissemination solution need to be located in the United States?	USPTO believes that the systems used to transform existing data sets for high volume dissemination need to be located in proximity to the existing data, most likely within the USPTO data center, because of the limited capability to move the data. The systems used to disseminate the data could be located anywhere.
21	What if any restrictions would be put in place on the internal or commercial use of data?	No restrictions other than timely and equitable delivery of the unaltered data provided by USPTO to the public at no cost.
22	Would a successful candidate be allowed to apply their brand to the public data dissemination platforms?	Yes. However, the raw data provided by USPTO to the successful candidate(s) must be labeled as such so that its authenticity is clear.
23	Would all users and commercial vendors have equal access to the data?	Published data must be delivered to all members of the public, including commercial vendors, within specified time frames in a manner that permits everyone to access the data at the same time without regard to geographic location. The requirements for universal delivery will be included in any subsequent procurement actions.
24	Would a successful candidate have any advantage over public or would there be equal access to the data and equal rights for its use, for all users and commercial vendors?	See the answer to question 23.
25	Can USPTO provide details of the budget allocated for Patent and Trademark Data Dissemination in FY 2009, and the budget proposed for FY 2010?	This information is not applicable to the market research being conducted at this time.
26	USPTO provides bulk data sets to commercial Patent and Trademark information companies and other organizations at marginal cost. Will USPTO continue the marginal cost policy if Data Dissemination operations are outsourced?	Yes. USPTO will continue to sell packaged data as long as there is a demand for that service.
27	What revenues did USPTO derive from Patent and Trademark Information Dissemination in FY 2008 and 2009?	This information is not applicable to the market research being conducted at this time.
28	Would USPTO be willing to provide facilities on the Campus free of charge for a Data Dissemination contractor/'partner?	Yes. The system(s) that is used to repackage the existing data must be hosted at the USPTO facility. The system(s) that are used to disseminate the data to the public may be hosted anywhere.

No.	Question	Answer
29	Liability - would a successful candidate be required to hold and save the Government, its officers, agents, and employees, harmless from liability of any nature or kind, including costs and expenses, for, or on account of infringement of any patent or copyright or any other unauthorized disclosure or use of any confidential secret, or proprietary data, process, product or invention, whether or not patentable, in the performance of an RFI related contract?	<p>Awardees must pass the data that USPTO provides to them on to the public, as is, at no charge. USPTO will retain liability for the accuracy of this data provided that it is unaltered prior to dissemination.</p> <p>The awardee will be solely accountable to it's customers for any data that is altered, repackaged, or modified in any way.</p>
30	Can the US Patent and Trademark Office make available the Road Map and Transformation Plan that was referenced in the FY2010 President's Budget Request?	This information is not applicable to the market research being conducted at this time.
31	Are there budgetary materials, perhaps a CIO's budget, which more clearly defines IT spending on various projects? While some of that was broken out in the FY2009 and FY2010 President's Budget Requests and the 2008 Performance and Accountability Report, more detail would be very helpful.	This information is not applicable to the market research being conducted at this time.
32	How many people work in IT at the USPTO and is this information possibly available broken down by type of position?	This information is not applicable to the market research being conducted at this time.
33	How does the USPTO distinguish, if at all, between data sets of a "bulk" format and data sets of a "machine-readable" format?	The term "bulk data" is used to represent a product that has been assembled as a collection and designed for data processing. "Machine-readable" data includes bulk data, but may also represent data that was designed for display but can also be read by a machine – for example, through Web-based "bots". The USPTO currently restricts machine readability on its public PAIR system through the use of a CAPTCHA mechanism that distinguishes between human- and machine-originating queries.
34	Are such data sets compliant with the USPTO Information Quality Guidelines?	Yes – all data sets are addressed under the USPTO Information Quality Guidelines.
35	Can the USPTO provide more details regarding the data format(s) used in and volume of each data set that is currently available in electronic format?	Yes. Attachment 1 to the RFI contains a broader description of the data being considered within the scope of this effort.

No.	Question	Answer
36	What, if any, data sets are not currently in electronic format?	<p>The data represented in Attachment 1 are all available in electronic format. The bulk of non-electronic data are the patent application file contents from 2002 and earlier. Other non-electronic data would be the data in Attachment 1 that does not fall within the identified date ranges.</p> <p>USPTO encourages awardees to convert paper or microfilm based records into electronic format if they believe that doing so provides a benefit to the awardee and/or the public.</p>
37	In what non-electronic format are such data sets maintained?	Not applicable – see answer above.
38	Would the Selected Vendor(s) be allowed to exclusively obtain and market these data sets?	Not applicable – see answer above.
39	The RFI appears to say that data production functions would remain with the USPTO, but the entire dissemination process would be moved to one or more Selected Vendor(s). Is this an accurate understanding of the RFI?	Yes.
40	In what form and with what frequency would the USPTO provide new data sets to the Selected Vendor(s)?	The intent is to meet or exceed the current data distribution timelines. Attachment 1 to the RFI contains the issue frequency of current data sets.
41	In what form and with what frequency would the USPTO provide changes to the data to the Selected Vendor(s)?	
42	What formats are "desired by the Intellectual Property (IP) community"?	PDF is a commonly-requested format that is not currently offered. Custom extracts of well-formed XML data are also desirable.
43	Is there an expectation that a Selected Vendor would standardize and make consistent bulk and machine-readable data formats that are now inconsistent across data sets?	Standardizing the current data is desirable as a component of preparing data sets for distribution in bulk. USPTO encourages awardees to address this in their responses.
44	Will a Selected Vendor be required to use currently used data formats (i.e., the Daily Application "C" file in XML for trademarks) and supply adequate advance notice to current recipients of data of any changes?	<p>See the answer to question 43.</p> <p>USPTO will retain responsibility for notifying recipients of changes to the data formats as it relates to the distribution of the raw data to the public.</p>
45	What are the USPTO's current demands from customers and related IT systems requirements to deliver the content? The RFI suggests that the USPTO does not have the necessary resources to accomplish its objectives. What, if any, budgetary estimates have led the USPTO to that conclusion?	<p>Current requests include unlimited data mining and electronic download of all data sets. USPTO's current infrastructure is incapable of supporting that volume.</p> <p>The budget estimates question is irrelevant to the market research being conducted under this RFI.</p>

No.	Question	Answer
46	What does the USPTO define as a "value-add"? For example, would offering the data sets on a high-speed connected server that allows the download in 1 hour rather than hours/days from the USPTO server be considered a value-add, or would value-add be defined as only data enhancements? Can the USPTO provide other examples of what the USPTO would consider to be value-added and what it would not?	At this point, USPTO is defining "value added" as data enhancements. Examples may include custom extracts of the data, custom combinations of the data to meet specific needs, or mash ups of data from multiple data sources. USPTO will consider other types of distributions enhancements in any subsequent procurement actions.
47	Would any restrictions be put on the Selected Vendor(s) in terms of when they could make the data available in its value added form (for which they would charge a fee) and/or when they can make the data available in a patent research platform for which they would charge a fee?	No restrictions other than the fundamental requirement for timely and equitable delivery of the unaltered data provided by USPTO to the public at no cost.
48	The RFI states that the Selected Vendor(s) must make bulk data available to the public at no charge. Would pricing of the value-added services have any government restrictions placed on it?	None are envisioned at this time
49	What advantages accrue to the benefit of Selected Vendor(s) in providing the services described in the RFI? Clearly, there are significant investments in staff/infrastructure, so there needs to be a way to recover this cost and run a margin.	Selected Vendor(s) may add value and market their value-added products.
50	What are the USPTO's initial expectations for Service Level Agreements and Quality Control?	See question 18.
51	Would the contract contain an "Authorization and Consent" clause under the Federal Acquisition Regulations? Or, would a Selected Vendor be required to hold the USPTO, its officers, agents, and employees, harmless from liability of any nature or kind, including costs and expenses, for, or on account of infringement of any patent or copyright or any other unauthorized disclosure or use of any confidential secret, or proprietary data, process, product or invention, whether or not patentable, in the performance of the contract?	See question 29.
52	Would updating the content with changes be a required activity that the Selected Vendor(s) would have to implement and offer for free (or would the Selected Vendor(s) simply be responsible for making the data provided by the USPTO available to the public in its original form and without comment or consideration for the actual content of the data)?	The vendor would be expected to maintain updates consistent with current USPTO practices, at a minimum. Withdrawn and corrected information is routinely made available to the public under current dissemination practices.

No.	Question	Answer
53	What specific activities does the USPTO see as being included in moving the entire dissemination process to one or more Selected Vendor(s)?	Availability of timely, free, downloadable bulk data at the customer's convenience is the essential activity.
54	How quickly after the data is made available to the Selected Vendor(s) would it have to be made available to the public?	The intent is to meet or exceed the current data distribution timelines. Attachment 1 to the RFI contains the issue frequency of current data sets. Specific time frame requirements for redistributing the data will be included in any subsequent procurement actions.
55	What customer support requirements would the USPTO envision the Selected Vendor(s) providing (e.g., only support for the technical downloading of the raw government data)?	Vendor support is expected to be limited to supporting the data delivery mechanisms. Support for the raw data would continue to be provided by the USPTO.
56	What customer support requirements does the USPTO envision they will maintain for this data? For example, Selected Vendor(s) will not be able to comment on the accuracy of the data. Will the USPTO maintain a resource to answer questions about the data that they provide?	The USPTO will provide support for the raw data content.
57	How will the USPTO react to the reporting of data errors either from the Selected Vendor(s) or from members of the general public? Will the USPTO maintain full responsibility for communicating to the public about data errors and plans for fixing the same?	<p>Awardees must pass the data that USPTO provides to them on to the public, as is, at no charge. USPTO will retain liability for the accuracy of this data provided that it is unaltered prior to dissemination.</p> <p>The awardee will be solely accountable to its customers for any data that is altered, repackaged, or modified in any way.</p>
58	Would parties other than a Selected Vendor have any restrictions on their ability to use, package or distribute data (e.g., in the form of search reports or via web services)?	None are envisioned at this time.
59	<p>Is there an expectation that a Selected Vendor would make applicable data available via distribution mechanisms similar to those mechanisms employed by the USPTO (e.g., HTTP)?</p> <p>Would controls exist to ensure a Selected Vendor could not make significant changes to distribution mechanism and procedures without advance notice?</p> <p>Would controls exist to ensure a Selected Vendor does not delay data dissemination to other providers or create a discernable competitive gap by making data either more complete or more current from their own systems?</p>	<p>The objective is to increase access through better distribution mechanisms.</p> <p>Appropriate controls will be evaluated after consideration of the information collected through this RFI.</p>

No.	Question	Answer
60	Would the USPTO be responsible for initially keying in new records, such that information pertaining to new records would be channeled to a Selected Vendor's infrastructure at the USPTO?	The USPTO will manage the source data.
61	Would the USPTO be responsible for electronically updating existing records, such that information pertaining to updated records would be channeled to a Selected Vendor's infrastructure at the USPTO?	The USPTO will manage the source data.
62	Would the USPTO be responsible for performing corrections to existing records and adding cross-indexing to records, such that all corrections would be channeled to a Selected Vendor's infrastructure at the USPTO?	The USPTO will manage the source data.
63	Is there an expectation that a Selected Vendor would make available in bulk or machine-readable format all existing information that is not currently in bulk or machine-readable format (e.g., all TDR information for trademarks)?	The Vendor would be expected to provide unaltered data for all data sets identified in Attachment 1 to this RFI.
64	Would the USPTO consider alternatives to the proposed in-house infrastructure solution, such as an infrastructure hosted completely offsite?	Yes.
65	Would the USPTO continue to maintain public interfaces for searchability (e.g., those available at the USPTO.GOV), independent of any systems that may be maintained by a Selected Vendor?	Yes.
66	Is there an expectation that a Selected Vendor would maintain all of the nearly two petabytes of data in the original formats currently maintained by the USPTO (e.g., as archives)? Is there an expectation that a Selected Vendor would make such data available in bulk or machine-readable formats? Or, might there be an understanding that this information could be maintained in the current format?	Information may be maintained and distributed via the most effective method for timely and equitable distribution, provided its content is unaltered from the authoritative data source.
67	Would there be any responsibilities for regularly contributing data to either the Trademark Official Gazette or the Patent Official Gazette?	No.
68	Would there be an option to implement an infrastructure solution only for trademarks, and not involving patents?	Not at this time.

No.	Question	Answer
69	The Obama Administration cites small businesses as important engines of economic recovery. Yet the ambitious size and complexity of the proposed endeavor precludes all but the largest Information Technology organizations from bidding successfully. Thomson, Reed, Google, Microsoft, IBM, and EDS readily come to mind. How does the Administration foresee small businesses and minority-owned businesses participating meaningfully in this development opportunity?	At this time an acquisition strategy has not been made. Teaming arrangements and subcontracting possibilities will be considered.
70	Much of the material submitted to the Patent Office is a matter of corporate confidentiality and even national security at some time in its life-cycle. What limits will Homeland Security place on vendors that have large proportions of foreign nationals in their employ?	Only published data will be available for dissemination. It is anticipated that any applicable IT or other security provisions will be included in any resulting contract.
71	The RFI contains a tacit admission that USPTO's Information Technology infrastructure is inadequate to the task of implementing the envisioned system effectively and is willing to surrender that responsibility. Yet it is axiomatic that the management and oversight of an Information Technology project requires substantially more expertise than the implementation tasks. How can the public and Intellectual Property community credibly trust that USPTO has the wisdom and capacity to oversee and manage this endeavor to a successful completion?	This comment is not specifically relevant to the market research currently being conducted.
72	The proposed scheme franchises Governmental Functions to the private sector in return for 'free' systems development for internal USPTO processes and some level of 'free' public access to the resulting database. The vendor, in turn, will profit by selling add-on features and data repackaging to its captive public audience. What USPTO mechanisms are in place to prevent the vendor from setting the 'free' baseline so low that the public will have to pay for everything beyond the trivial and inadequate data access the USPTO presently provides?	The requirement is for timely, equitable, and no-cost access to USPTO data in bulk.
73	The RFI depicts a hypothetical outsourcing proposal that simultaneously conserves agency resources, promotes free public access to public data, and provides adequate profit margins to the private sector. Can USPTO point to a precedent for a completed project of this nature and magnitude that optimally balances these competing interests?	Although we cannot point to a project of identical nature and magnitude, the USPTO is aware of other Federal projects whose aim was to provide free access to public data at no cost to the Government while providing commercial opportunities for the private sector. The Government Printing Office recently announced a contract opportunity for the digitization of documents at no cost to the government.

No.	Question	Answer
74	What development time-frame does USPTO envision for procurement, implementation, and operational life? From the ambitious scope, this looks to be a five-year project at very least. Does USPTO intend to suspend needful improvements to its existing prepublication platforms in deference to this new enterprise?	An implementation timeframe has not been established. Responders to the RFI are encouraged to provide thoughts on the timeframe for implementing their proposed solutions.
75	Federal Information Technology procurements have a sad history of being overweight at birth, obese during midlife, and sclerotic in old age. Some have been obsolete even before completion – FBI's systems come readily to mind. Surely USPTO recognizes the problem since some of its own systems are in a state of legacy paralysis. They are so ponderous and baroque that only the incumbent contractor can successfully bid on maintenance. What novel design approach does USPTO intend to employ to insure that this effort does not go similarly awry?	This comment is not specifically relevant to the market research currently being conducted.
76	The RFI mentions mostly prepublication and assignee data but is silent regarding the Application and Grant Red Book and Yellow Book data which has been routinely packaged and sold in bulk by USPTO for over 30 years. This is likewise public data. Is a similar scheme anticipated by USPTO for its dissemination? Most of its users would readily state that it is priced too high relative to other Patent Authorities such as EPO, JPO, WIPO and SIPO. Yet these users do not uniformly embrace 'free' distribution for fear data quality will suffer. Has USPTO studied the impact such a change would have on these ongoing activities?	<p>The Red Book and Yellow Book data is included in this RFI, and is described in Attachment 1.</p> <p>In accordance with the President's Open Government Initiative, the intent of this RFI is to make data more accessible.</p>
77	Is the data all electronic?	Yes.
78	If there is physical data, is there an expectation to convert the data to electronic within the scope of this project	Not applicable under this RFI.
79	Can you clarify what you mean by machine readable format?	See answer to number 33.
80	<p>The RFI describes an estimated 2 petabytes of data sets maintained by USPTO. Could more detail about each of these data sets be provided such as:</p> <ul style="list-style-type: none"> a. Format(s) b. Number of records/pages/documents c. Estimated size of each data set 	Yes. Attachment 1 to the RFI contains a broader description of the data being considered within the scope of this effort.
81	Could sample documents of each data set be provided?	USPTO will consider providing example data sets with any subsequent procurement action.

No.	Question	Answer
82	Is it a requirement that any potential solution be housed at USPTO offices? Could a vendor supplied location be used instead?	USPTO believes that the systems used to transform existing data sets for high volume dissemination need to be located in proximity to the existing data, most likely within the USPTO data center, because of the limited capability to move the data. The systems used to disseminate the data could be located anywhere.
83	Is the hardware and software to be mandated by the USPTO or does the vendor have the ability to provide that as part of a proposed response?	Responses should propose the tools that will be part of the solution.
84	How does USPTO envision that the data to be disseminated at no charge be handled? Would dissemination via the web and / or FTP be sufficient for the USPTO?	Responses should describe any appropriate dissemination methods that ensure timely and equitable access to data.