

**From:** Mark DesMeules [email redacted]

**Sent:** Wednesday, May 06, 2015 4:14 PM

**To:** WorldClassPatentQuality

**Subject:** Comments of the Coalition for Patent and Trademark Information Dissemination

To whom it may concern,

Please find attached the comments of the Coalition for Patent and Trademark Information Dissemination (CPTID) in response to the USPTO's request for: comments/suggestions regarding patent quality.

Thank you,

Mark DesMeules – on behalf of CPTID

**Mark H. DesMeules**



American Continental Group  
900 19th Street NW  
8th Floor  
Washington, DC 20006  
(202) 327-8131

# Coalition for Patent and Trademark Information Dissemination

May 6, 2015

Michelle Lee

Under Secretary of Commerce for Intellectual Property and Director of the United States Patent and Trademark Office  
600 Dulany Street  
Alexandria, VA 22314

RE: Comments/suggestions regarding patent quality

Dear Under Secretary Lee:

We appreciate the opportunity to express views on behalf of the Coalition for Patent and Trademark Information Dissemination (CPTID) on the USPTO's recently announced Quality Initiative. The CPTID is a group of entities committed to the notion that private sector participation is essential to the quality and integrity of the US patent and trademark system. We have a significant interest in the quality of the information made publicly available by the USPTO and strongly believe that the US patent and trademark system depends not only on the quality of the raw data disseminated by the USPTO, but also on the value-added information, tools and services provided by private sector entities such as our members.

In introducing its new quality initiative, the USPTO has quite correctly said that “[e]ffectively promoting...innovation requires that issued patents fully comply with all statutory requirements and, of equal importance, that the patent examination process advance quickly, transparently and accurately.” The agency’s goal is “[h]igh quality patents” that “permit certainty and clarity of rights, which in turn fuels innovation and reduces needless litigation.” Such a goal is laudable and deserves to be applauded by all stakeholders. However, the issue of patent quality has many dimensions and those dimensions have varying degrees of importance depending upon one’s perspective. The most obvious dimension of quality relates specifically to the accuracy or correctness of patent examiners’ decisions, but this is by no means the only dimension of quality. For example, consistency of the examination process across examiners in an art unit, and across art units, is an important measure of quality to applicants and patent attorneys. And accuracy and comprehensiveness of published patent content is highly important to users of public patent information and is very relevant to the USPTO’s Pillars I (Excellence in work products) and II (Excellence in measuring patent quality). This document will specifically address how the CPTID believes the USPTO should address accuracy and comprehensiveness in Proposals #2 (Automated Pre-Examination Search) and #3 (Clarity of the Record) in Pillar I and Proposal #4 (Review of and Improvements to Quality Metrics) in Pillar II.

## **Pillar I Excellence in Work Products, Proposal #2 Automated Pre-Examination Search**

The USPTO has proposed offering examiners improved automated-pre-examination search of prior art to increase efficiency and quality of patent examination. While the CPTID understands the intentions of the USPTO in this effort, it does not support the proposal. The implementation of such searches for examiners could result in too heavy a reliance on the automated results. The CPTID believes that an automated pre-examination search is no substitute for an examiner’s own, more thoughtful and informed manual searches (assisted by advanced search tools). An

## Coalition for Patent and Trademark Information Dissemination

examiner reliance on automated pre-examination searches could result in relevant results being missed that only a search with human touch can produce. As a result, such searches may lead to decreased patent quality, thus having a detrimental effect on the USPTO's overall patent quality initiative. The resulting decrease in patent quality would introduce unwanted uncertainty into the innovation environment. The CPTID recommends that the USPTO continue to offer examiners a suite of manual search tools that facilitate the examination process and thus the issuance of high quality patents.

If, however the USPTO decides to move forward with Proposal 2: Automated Pre-Examination Search, the CPTID has several recommendations that would reduce the chance that such an offering would lessen the quality of patents issued by the agency. The CPTID would recommend that, rather than develop its own automated search tools in-house, the USPTO should use the resources of its private sector partners; technology for such advanced search already exist and can be integrated with the USPTO's existing examiner tools. The use of honed, private sector technology would ensure the efficacy and consistency of automated pre-examination search results. The CPTID would also advise the USPTO to ensure that any search tools it employs place an elevated emphasis on data quality and clarity, thus facilitating the production of search results that are of the highest level of usability to examiners. If the agency decides to move forward with this proposal, the USPTO would likely need to provide the tool to the applicant as well – requiring licenses to the commercial databases in use at USPTO.

### **Pillar I Excellence in Work Products, Proposal #3 Clarity of the Record & Pillar II Excellence in Measuring Patent Quality, Proposal #4 Review of and Improvements to Quality Metrics**

The remainder of the document will discuss six aspects of quality that can easily be overlooked when considering excellence in work products and excellence in measuring patent quality but are of utmost importance to the long-run integrity and quality of the US patent system.

**1. The Quality of Inputs to Examination** - The quality, suitability and accuracy of the information provided by applicants to the USPTO have direct bearing on an examiner's activity and consequently the outcome of the examination process. Analysis of applicant-submitted documents to the USPTO shows that applicants make literally millions of procedural, categorization, and other errors in their submissions to the USPTO each year. Currently the USPTO has contractor-operated processes in place that filter and correct documentation submitted by applicants. This includes, for example, correcting applicant-supplied doccodes, which in turn enables examiners to efficiently locate documents contained in the file wrapper. This filtering is critical to effective and efficient work on the part of patent examiners. New e-filing technologies and methodologies – such as Text2PTO – will not solve this problem, because the vast majority of applicants file fewer than ten applications per year, far too few to have real command of USPTO requirements and business rules.

In addition to maintaining existing quality checking on applicant inputs, filtering for applicant-generated errors can and should be expanded and enhanced by using advanced text analytics to identify structural flaws and other document shortcomings. A wide range of anomalies and error conditions can be detected through automation, and the USPTO should be applying such technologies before the start of examination, thus creating opportunity to improve efficiency during examination. Such tools are commercially available in the marketplace today, including

## Coalition for Patent and Trademark Information Dissemination

from our members. Relatively straightforward examples of common detectable errors include claim numbering errors (particularly when claims are revised by inserting or deleting whole claims) and lack of antecedent basis. Many other examples exist. Identifying such problems up front could provide significant benefits to patent examiners and improve the efficiency and quality of examination. The USPTO should explore technologies that commercial entities have already built that perform such analysis and implement tools/processes that perform these functions.

### **2. The Quality of the Examination Process in Relation to Structural/Procedural**

**Accuracy** - The examination process, which commonly takes place over the course of multiple months (sometimes years), may involve replacement of the application specifications or entire claim sets. The opportunity to introduce structural errors becomes a much greater risk with each submission. Review of allowed patent applications shows that well over 100,000 flaws relating to policy or statutory requirements and/or agency business rules persist up to the point of allowance each year. These are identified post-allowance today by independent review in the publication process – a key pre-issuance quality filter. It is critical to the quality of published USPTO content that this quality check and correction step be retained. If it is not, the clarity of the record will be diminished and users of public patent information will be negatively impacted. Additionally, technology exists to apply text analytics upon introduction of new content as well as at the point of allowance that could help to avoid the persistence of an additional range of structural errors in allowed patents. The USPTO should keep existing post-allowance filters in place, but in addition to those filters it should also implement text analytics that are already available in the private sector to catch errors and anomalies in the examination cycle with a view to further reducing the persistence of errors to the point of allowance or beyond.

**3. Consistency of Performance across Examiners and Art Units** - Currently a very wide range of examiner performance exists both within individual art units and across art units. This variation in performance is often overlooked as a quality issue, but it can be a source of significant frustration for patent applicants and a cause (as well as a symptom) of inefficiency in the patent prosecution/examination cycle. Automated tools exist that quantify such variation, and quantification is a necessary component of an overall improvement plan. For example, technology exists in the private sector that can predict with a high degree of accuracy how long it will take for an RCE to be considered on an Art-Unit-by-Art-Unit basis, or what the likelihood of allowance is on an examiner-by-examiner basis. Such tools are already beginning to change applicant behavior. The USPTO should consider implementing such tools within the agency in an effort to reduce unwarranted and undesirable variation and improve the quality of applicant experience as well as the efficiency of the examination process.

**4. Accuracy of Searchable Published Patent Content** - Accuracy of the text and other searchable components of published patent documents is absolutely fundamental to the clarity and accessibility of the public record. This is especially true in a data set as large as the entirety of published U.S. patent content. It is easy to take this for granted, but it would be a major mistake to do so. An extraordinarily large number of documents are processed by the USPTO and USPTO contractors today using Optical Character Recognition (OCR) technologies. OCR'd text is known to contain a significant number of errors. Human review overcomes this problem by correcting an estimated 200,000,000 text errors each year in OCR'd patent documents prior to publication. These errors are corrected by highly skilled human proofreaders rather than relying

## Coalition for Patent and Trademark Information Dissemination

exclusively on automated correction processes. Such automated processes alone would not produce text that is accurate enough for patent research. Algorithms and spelling dictionaries used by automated processes are notoriously lagging behind the technology fields represented by patents. These automated processes alone would inevitably introduce deviations from the patent source documents. Deviations cannot be allowed because the application submitted must be faithfully recorded. Failure to correct OCR errors would have immense negative consequences for all stakeholders who rely on the accuracy of this data. The USPTO's Patent Application Text Initiative (PATI) – which relies on OCR technology alone – produces outputs that are sufficient for some purposes, but not for publication, and not for the detailed analysis applied to such data by commercial entities and researchers.

Commercial providers add value by enhancing raw data to create new features and functionality. An example of added value is LexisNexis's Semantic Search feature. This feature analyzes a phrase or a paragraph of text entered by the searcher. Behind the scenes, it makes intelligent connections to other words and phrases that may or may not have occurred to the searcher. After analysis of the phrase or paragraph, Semantic Search offers the searcher additional terms and phrases it identified as related. The intelligent connections made by Semantic Search are supported by a massive database that was created by a proprietary algorithm. The algorithm examined the text of millions of USPTO patents and Elsevier scientific journals to find connected terms and phrases. It then stored the connected terms in a database. This database is used to provide the searcher with a more intelligent search experience. Imagine the impact that USPTO raw data with billions of OCR errors (200,000,000 per year over multiple years) would have on a tool of this sort. ProQuest also has a "more like this" functionality which extends both to Patents and Non-Patent Literature. ProQuest also has a patented process for indexing and searching graphical materials. Today that product looks at non-Patent literature but could be extended to drawings in Patents. In addition, Thomson Reuters offers a patented "Smart Search" tool that leverages the significant quality in the Derwent Worlds Patent Index. The Derwent World Patents Index database is created by worldwide teams of multilingual scientific and technological experts that correct and clarify the key novelty aspects for each patent family. The proprietary "Smart Search" uses the world's best Semantic search technology to execute patent searches using human-enhanced, quality controlled knowledge built into the enhanced data fields of the Derwent database.

The USPTO should maintain its current contractor-operated text correction processes and patent text quality checks to ensure the continuing high quality of published patent content at 99.995% accuracy rates or better.

**5. Richness of Metadata / Depth of Content Tagging** - U.S. patent content today has a very rich layer of metadata that facilitates utilization of the content for a wide range of purposes across the IP stakeholder community. The richness of this metadata is itself a significant measure of quality. The USPTO should not consider reducing the richness of the metadata produced today as a tradeoff to achieve higher degrees of automation and lower data capture / content processing costs. Nor should it do so because of a desire to shift the responsibility for tagging to applicants, knowing that applicants will often get the tags wrong if the requirements are too complex. Reducing the richness of the tagging for any such reason would be a significant mistake. Users of patent information in the US and around the world rely on the rich metadata produced today by the USPTO and consider it the gold standard for published patent information

## Coalition for Patent and Trademark Information Dissemination

globally. The USPTO should not consider any simplification of current metadata / tagging requirements. On the contrary, enriching the metadata further – for example, to allow linking of part numbers to drawings or to cross reference different chemical names for identical chemical entities – would increase the value and perceived quality of U.S. patent content, and would foster even greater innovation. Many commercial database providers also have extensive indexing schema attached to their products. The Institution of Engineering and Technology's Inspec database, for example, indexes non-patent literature with International Patent Classification codes. ProQuest Dialog users rely heavily on the metadata and indexing to sort vast quantities of content and also hone in on search minutia. Thomson Reuter's Thomson Innovation platform integrates the International Patent Classification codes within the Inspec database into the intellectual property search and analysis environment as a key prior art resource and non-patent literature analysis component.

**6. Consistency of Published Content** – Patent applications are filed in the U.S. by a wide and diverse range of over 30,000 patent agents / attorneys. In combination with the novel nature of the content contained in patent applications, this fact would yield significant variation in the way similar material is presented to the public (e.g. in the way in which tables or chemical structures are published) without a processing step to introduce consistency across all applications. The agency currently utilizes a contractor to ensure such consistency. Failure to continue this practice would likely result in a reduction in perceived quality of published content across the stakeholder community, and would certainly result in an inferior product compared to today's published patent products. Therefore the USPTO should retain this processing step prior to publication.

Accuracy, completeness and accessibility of the public record, and consistency in applicant experience across patent examiners and art units are important dimensions of quality that must not be overlooked as the USPTO defines and pursues its quality objectives. These dimensions can easily be taken for granted, but this would be a significant mistake as they are essential to an efficient patent and trademark system that is vital to US innovation, economic growth and the public good. The agency should explore each one thoroughly and ensure that any changes to current processes that are contemplated will indeed enhance the quality (and the perception of the quality by stakeholders) of USPTO processes and published content, and not diminish it.

We hope that the above points assist the USPTO in taking into account the wide range of issues that relate to the quality initiative the agency has undertaken. We also hope the USPTO will indeed take these points into consideration as it makes decisions about how to make the best patent system in the world even better.

Thank you,

Marla Grossman  
Executive Director  
Coalition for Patent and Trademark Information Dissemination