

Impact Statement Proposing Improved Open Data Accessibility

January 1, 2020

Andrei Iancu
Under Secretary of Commerce for Intellectual Property
Director of the United States Patent and Trademark Office

USPTO Madison Building
600 Dulany Street
Alexandria , VA 22314

Dear Director Iancu,

The USPTO should be commended for being at the leading edge of electronic access to a Federal agency. The PAIR and EFS access is beyond what almost any Federal agency has envisioned. By providing free access to Public PAIR, the patent community has evolved to automate many of the tasks needed to protect inventors' interests and evaluate the landscape of IP data to support more efficient business endeavors and decisions.

The bargain of the patent system memorialized in the Constitution is to provide a limited exclusivity to innovators if they fully disclose their inventions to the public. In other words, the primary benefit to the U.S. public is derived from the details of the U.S. patent documents. Now,

the number of U.S. patents exceeds 10 million, and the U.S. public needs help to derive important benefits from this large pool of documents.

The public is able to draw a greater benefit when the Office provides IP data as broadly as possible, allowing interested parties to investigate and build upon prior innovation. Without full and open access to the documents in the prosecution history, information crucial to this goal is difficult to obtain, resulting in confusion such as an inability to confidently define a patent's claim scope. As has been broadly discussed in the ever-changing landscape of patent validity, uncertainty breeds hesitancy as innovators decide if, when, and where to file patent applications.

With recent changes to PAIR and load-related stability issues resulting in reduced accessibility of prosecution histories, patent costs and quality are suffering. Many docketing and IDS tools no longer function so that error-prone manual processing is required. EFS and the payment system is unreliable enough such that staff is trained to revert to paper filings with costly, late-night trips to the post office followed by days of worrying until the filing can finally be verified in PAIR. Also, many electronic filings crash in the process such that it is not clear if the filing happened and/or the fee was paid.

A group of stakeholders has organized to provide suggestions on increasing reliability of the public facing systems while furthering the stated policy goals of disseminating IP data. To improve reliability of these systems, there is an immediate need to fix PAIR access while also decreasing the load on the servers that host PAIR, EFS or other USPTO services. We have formulated our suggestions into immediate, short term and long term proposals.

Immediate Action:

MyUSPTO users could previously access both Private and Public PAIR from one browser session. A few weeks back, that was disabled. Accessing file wrappers not associated with a user's customer number now requires the user to navigate to public PAIR in a new browser instance and successfully complete a CAPTCHA test. The user then has to keep both the private PAIR and public PAIR sessions active through time-out algorithms in two browsers or be logged out and have to start all over. For users, this is a huge waste of time. For automated docketing and IDS solutions, this requires large-scale rewriting of software. And once the rewrites are complete, the result will simply be more load on

public PAIR (a system which was already prone to being unusable with common “high system volume” or other errors).

Our proposal is to immediately roll-back the change so that credentialed access through MyUSPTO enables the logged-in user access to file wrappers that are not associated with the user’s customer number. To reduce load on the Office infrastructure, it is recommended that only licensed users or their delegates get access via MyUSPTO. Should there be inappropriate scraping, the OED has jurisdiction over licensed practitioners and may provide clear instructions, warnings, and even issue suspensions.

Short-term Action:

Provide a copy of Public PAIR file wrappers to further the stated Open Data goals to disseminate IP data. Stakeholder tools depend on this data to provide the functionality users demand. It is appreciated that a tragedy of the commons was created by the Office’s leadership in making Public PAIR freely available. Many automated processes individually gather this information to create a collective burden on Office infrastructure that causes instability. With the ease of automated data gathering today, any free interface to desired data will have stability problems. The Office is a victim of their success. Although portions of this information have been hosted by third parties for free in the past, the information is incomplete and/or difficult to access effectively.

We propose that the Office facilitate inexpensive transfer of Public PAIR data in a way that makes “free” scraping unattractive. The signatories to this Impact Statement offer to provide cloud hosting of an initial copy of Public PAIR (“snapshot”) that would enjoy regular daily updates (“deltas”). That copy can be provided very inexpensively to the stakeholders in a manner far cheaper than any automated scraping effort. The Office can provide this in any format convenient and use existing protocols and infrastructure, for example, the methods used for gathering the Office Actions Dataset and the Office Actions Rejections and Citations APIs available at developer.uspto.gov. If a hard drive or tape format or particular cloud host is preferred, the Open PAIR Coalition will support the lightest touch on PTO infrastructure and staff. In any event, the need is so great in the stakeholder community that any sort of accommodation can be figured out by the Coalition.

Long-Term Action:

Although third-party cloud hosting of PAIR data can quickly undermine any incentive to scraping Public PAIR, free dissemination of IP data is a stated policy priority of the Office. The Patent Examination Data System (PEDS) along with the Open Data Portal already provide metadata on Public PAIR to stakeholders through APIs. However, full access to the prosecution histories and richer portions of the application have not been provided by API. Additionally, PEDS and other data is not updated at the frequency required by stakeholders. The absence of this IP data has largely created this tragedy of the commons.

It is proposed to finish PEDS or provide another API to Public PAIR information with further input from the stakeholder community. Source documents, XML or other formats and metadata already in the Office systems should be provided where available (e.g., status indicators, current claims, search queries, etc.) with emphasis on more information even if the format is not perfect so that the stakeholders can leverage the pre-processing already performed using their fees. Data in this system should be updated in real time (like PAIR) so that there is not an incentive to burden the servers that host PAIR & EFS with massive numbers of redundant requests for data to get it quicker as the information is, after all, public. Arrangements can be reached with stakeholders to further reduce the burden on the active systems, such as mirroring portions of the data on a third-party host. As a result, the public would have access to not only basic patent metadata (including dates, statuses, titles, inventors, and applicants), but the rich prosecution histories (arguments in office actions, responses, ex parte appeals, etc.) and the wealth of knowledge in the separate portions (specifications, claims, drawings) of the applications themselves without needing to pull each document separately.

Members of the public should not need to form a tech company in order to gain the basic insights that this long-term proposal provides, and making this information more accessible will allow the researchers and tech companies in this space to provide new insights (e.g., better analytics, search algorithms, and automation tools) that take the file history into account to increase the overall quality and efficiency of the patent system. If the U.S. Patent Office continues to embrace innovation in the patent space, these new insights will push the U.S. Patent Office to even greater heights relative to peer offices.

In conclusion, the stakeholders wish to commend the Office on their Open Data efforts thus far. Recent changes and instability of these systems has caused extra expense and lower quality patent prosecution.

Best regards,

The Open PAIR Coalition

Coalition@OpenPAIR.org

Thomas D. Franklin	CEO	Triangle IP
Chad M. Gilles	CEO	BigPatentData
David Nickelson	Dir. of Research	Techson IP
Ian Schick	CEO	Specifio
Chris Holt	CEO	IPDataLab
Jim Goedken	CEO	SyncIDS
Russ Krajec	CEO	Bluelron IP
Peter Kramer	Principal	Ingenium Patents, LLC
Rocky Berndsen	Co-founder	PatentPrufer, LLP
John Harrity	Managing Partner	Harrity & Harrity, LLP
Daniel Tesler	Principal	Law Offices of Daniel A. Tesler LLC
Jeff O'Neill	Principal	Patent Bots LLC
Curtis Wadsworth	CEO	Three10 Solutions, Inc.
Jonathan Jaech	Partner	One LLP
Ury Fischer	Shareholder	Lott and Fischer PL
Richard Jefferson	CEO	Cambia & Lens.org

Amy Pearson	Managing Partner	TenacityIP LLC
Ido Tuchman	Founder	Busylabs
Chakshu Kalra	Director	GreyB Services Pte. Lte.
Ben Dugan	Founder	Lawtomata LLC
Thomas Marlow	CTO	Black Hills IP

cc: Thomas Beach, Chief Data Strategist and Portfolio Manager at USPTO
Damian Porcari, Director of the Elijah J. McCoy Midwest Regional USPTO