MEMORANDUM ON THE STUDY OF DIVERSITY AMONG PATENT APPLICANTS

I. Background

This memorandum presents the conclusions of the Diversity of Applicants initiative required by the Leahy-Smith America Invents Act. Section 29 of the AIA directs the Director of the U.S. Patent and Trademark Office (USPTO) to "establish methods for studying the diversity of patent applicants, including those applicants who are minorities, women, or veterans."[1]

Section 29 required the USPTO to fulfill the statutory provision by March 16, 2012. The USPTO timely established and published its methodology by that date.[2] The USPTO has since implemented its established methodology, which proceeded in three parts.

First, the USPTO shared public patent data with the Census Bureau's Center for Economic Studies (CES), so that CES could match this data with diversity information at the Census Bureau. To comply with the Paperwork Reduction Act,[3] Privacy Act,[4] and the Census Bureau's confidentiality obligations,[5] CES provided only tabulations of aggregated information about the diversity of patent applicants, with no information on any individual applicant.[6] CES also advised the USPTO on the accuracy and reliability of these statistics, including potential biases from any inability to match the data adequately.

Second, the USPTO issued a *Federal Register* notice seeking public comment on the Diversity of Applicants initiative.[7] The notice asked whether and how to study patent applicant diversity further, including the value of a survey and other data collection options, the use of personal identifying information, the sharing of data with other institutions, and the assurance of data accuracy.

[1] Pub. L. No. 112-29 § 29 (Sept. 16, 2011).
[2] United States Patent and Trademark Office, *Diversity of Applicant Methodology* (Mar. 16, 2012), *available at* www.uspto.gov/aia_implementation/programs.jsp.
[3] 44 U.S.C. § 3501 et seq.
[4] 5 U.S.C. § 552a et seq.
[5] 13 U.S.C. §§ 9, 214.
[6] The highly aggregated group data that CES provided to USPTO was devoid of any personal identifying information. Because sensitive Census information concerning diversity characteristics is protected under Title 13 of the United States Code, once USPTO information became commingled with Census data, that commingled data was also confidential and could not be released.
[7] Request for Comments on Methods for Studying the Diversity of Patent Applicants, 78 Fed. Reg. 72064 (Dec. 2, 2013).

Third, based on tabulations and guidance from CES and on comments from the public, the USPTO has reached a determination about whether, and under what restrictions, to take further steps toward describing patent applicant diversity accurately.

II. Implementation and Findings

A. Data-Matching with the Census Bureau

To analyze currently available information about patent applicant diversity, the USPTO shared its publicly available data about patent applications with CES. Consistent with the language and legislative history of Section 29 of the AIA, the analysis sought to describe two things:

(1) the overall, cumulative (i.e., highly aggregated) demographic characteristics, such as race, gender, age, and geography, of inventors as a group; and
(2) the overall, cumulative (i.e., highly aggregated) business characteristics, such as revenues, number of employees, and geography, for companies as a group.[8]

Importantly, this analysis gathered and evaluated cumulative data only on groups of individuals and companies, not data that would identify any particular individual or company. Moreover, by using existing data and cooperating with CES, the USPTO did not impose any additional burden on patent applicants while protecting the identity of particular individuals and companies.

The data that the USPTO provided to CES for this analysis consisted only of certain public information provided on the face of patents granted between January 1, 2005, and December 31, 2006. Specifically, this information was the name and address of the inventor, generally only the town and state. CES then confidentially attempted to match this data against its own data with the goal of identifying the cumulative demographic information of the inventors as a group.

The analysis was only partially successful. CES was able to match 64.3% of the U.S.-resident inventors provided by the USPTO. The basic information that the USPTO had collected from inventors—i.e., name, town, and state—was not a strong basis for matching with Census data. For example, it was usually not possible to match common names (such as "John Smith" or

[8] The intent of Sec. 29, to study the diversity of applicants, reflects a broad understanding of USPTO "applicants" as including organizations as well as individuals. Indeed, Rep. Gwen Moore, the sponsor of Sec. 29, specifically refers to businesses owned by minorities and women as consumers of the patent process, i.e., applicants:

> I certainly do applaud USPTO for their outreach to the Women's Chamber of Commerce and to the National Minority Enterprise Development Conferences to try to increase diversity with utilizing the patent process. But some recent data have raised concern that minorities and women-owned businesses are just not keeping up with the patent process.

157 CONG. REC. H4484 (daily ed. June 23, 2011) (statement of Rep. Moore) (emphasis added). Rep. Moore's discussion continues to refer interchangeably to firms and individuals as applicants whose diversity is of interest:

> Preliminary data from a 2009 Kauffman Foundation survey of new businesses show that minority-owned technology companies hold fewer patents and copyrights after the fifth year of starting than comparable nonminority businesses. In fact, the Kauffman data show that minority-owned firms with patents hold only two on average, compared with the eight of their counterparts. Another survey uses National Science Foundation data to suggest that women commercialize their patents 7 percent less than their male counterparts.

Id. (emphasis added).

"Mary Johnson") in large cities (such as "New York, NY" or "Chicago, IL"). The poor quality of data-matching and some statistical bias suggest that the limited information that the USPTO currently collects about inventors is not sufficient to allow CES meaningfully to describe the cumulative diversity characteristics of inventors as a group within the meaning of Section 29.

B. Public Comments on Further Studying Diversity

Following the partial success of this cooperative effort to match the USPTO's publicly available patent data and the Census Bureau's confidential demographic data, the USPTO sought public comments on whether or how to collect further information pursuant to the *Diversity of Applicant Methodology*. The USPTO received four comments: one from the American Intellectual Property Law Association (AIPLA) and three from individual members of the public.[9]

The comments largely supported the general idea of further studying patent applicant diversity,[10] though this view was not unanimous.[11] Those who supported further study generally favored the use of surveys to collect information.[12] One commentor further suggested that the USPTO repeat its survey on an ongoing basis to monitor changes over time.[13] As for gathering data from other institutions or organizations, the AIPLA expressed confidence that the USPTO can effectively do so with respect to U.S. applicants and corporations, but cautioned that this may be more difficult in the international context.[14]

Notably, the AIPLA urged in strong terms that the USPTO should conduct any surveys only on a voluntary basis,[15] and another commnentor agreed that voluntary surveys would reassure respondents about their privacy.[16] Legitimate privacy concerns notwithstanding, however, none of the comments addressed concerns of statistical validity that collecting data on a voluntary basis would likely affect the accuracy of the information that the USPTO would receive.

III. Findings of the USPTO

The USPTO has determined that the ability of mandatory surveys to generate individual demographic diversity data of acceptable quality and reliability is in tension with the lack of public support for mandatory surveys due to privacy concerns under current law. There is support in the public comments for voluntary surveys, but there is also no indication in the public comments that information collected through voluntary surveys would be of comparably adequate accuracy as would information collected through mandatory surveys. Conversely, there is reason to expect that voluntary responses would contain statistical bias arising from self-selection among respondents. This risk for statistical bias is particularly relevant because

[9] United States Patent and Trademark Office, Comments on Methods for Studying the Diversity of Patent Applicants, *available at* www.uspto.gov/patents/law/comments/applicant_diversity_comments.jsp.
[10] Comments of AIPLA at 2; Comments of Nickolaus Leggett at 1; Comments of David Martinez at 1–2.
[11] Comments of Glenn Johnston (rejecting any governmental study of diversity in the innovation system and suggesting that privately funded research would be the appropriate vehicle for studying such diversity).
[12] Comments of AIPLA at 2–3; Comments of Nickolaus Leggett at 1; Comments of David Martinez at 2.
[13] Comments of Nickolaus Leggett at 1.
[14] Comments of AIPLA at 3.
[15] Comments of AIPLA at 2.
[16] Comments of David Martinez at 2.

the initial findings of USPTO-CES data matching effort were of limited value due to statistical distortions from inadequate matching.

In order for the USPTO to study patent applicant diversity further, there must first be a resolution to the tension under current law between the statistical rigor of mandatory surveys and the public support and existing authority for voluntary surveys. Until such a resolution, the USPTO notes that parallel efforts are underway to study demographic, economic, and other forms of diversity in the innovation system. A prominent example within the Federal Government is the Census Bureau's *Business Dynamics of Innovating Firms: Linking U.S. Patent Data with Administrative Data on Workers and Firms* working paper.[17] Examples from the academic community include the *Understanding the Demography of American Innovation* research initiative of the Information Technology & Innovation Foundation and George Mason University as well as the *Empowering Women in Technology Startups* initiative of the University of Florida Innovation Hub.

IV. Conclusion

As expressed in the original *Diversity of Applicant Methodology*, the USPTO remains committed to responding adequately to the concerns of Congress and the increasing need in the innovation economy to analyze and understand how the various parts of the Federal Government are responding to the needs of innovators. The USPTO's approach to evaluating the collection of potentially sensitive information from patent applicants has been iterative and careful in order to ensure that the USPTO continues to be respectful both of current laws and regulations and of important concerns expressed through the USPTO's ongoing engagement with the public.

[17] Stuart Graham, Cheryl Grim, Tariqul Islam, Alan Marco & Javier Miranda, *Business Dynamics of Innovating Firms: Linking U.S. Patent Data with Administrative Data on Workers and Firms*, Census Bureau Working Paper (forthcoming 2015).