

## Event

Policy makers and academics use patent data to formulate innovation and patent policies.

## Significance

While patent data provides a rich mine of empirical information, it is often taken as the sole measure of innovation. Given differences in patent law between countries, strategic patent filings and differences in reliance on patents in different fields of technology, patents are a poor measure, standing alone, of actual levels of innovation. This leads to the important challenge of not only better collecting patent data, but of contextualizing results through a broader set of empirical measures.

## Analysis

Patent data are used by researchers as a proxy measure for knowledge and employed to probe multiple aspects of innovation. Taken as outputs these can provide insight on the effect of changing patent law or policy, while as inputs they can indicate whether and how patents affect growth, performance, revenue, welfare and follow-on innovation. 'Gene' patent data highlights many of the inherent challenges. At present, global patent offices employ disparate practices with discrepancies in patent and claim formats and identifiers across patent office and biological databases, and delayed listing of claimed sequences. While recent efforts such as Cambia's Patentseq and the Lens are providing free, public means to search and analyze patents including sequence-based inventions, access to complete data across jurisdictions is still sorely lacking. Coordinated international efforts toward standardization, alignment, and providing up-to-date records in accessible electronic formats is a pressing imperative and will allow for cross-jurisdiction analyses.

Flawed analyses can lead to mistaken conclusions and have far-reaching consequences when relied upon as the basis for policy or law-making. Patent analyses often rely on stratification by the type of invention claimed or by the way specific types of claims are made. Employing algorithms to categorize claims can be time and cost-efficient for analysis of large-scale data. However, they can be problematic because of variation in language and formulation, and the heterogeneity of patent law across jurisdictions. While hand-analyses can limit these problems, it is expensive, may be unworkable in large datasets, and is vulnerable to reliability concerns.

Deep understanding of the relationship of patents to innovation will require examining all stages of R&D commercialization pathways including endpoints measuring welfare and how revenues from patents are reinvested by assignees. New and creative ways to analyze claims, better ways of tracking the people involved in the creation and flow of knowledge through innovation pathways, access to rich datasets held privately or subject to high proprietary barriers (for example, licensing data), and better indicators to measure effects on public good are all urgently needed. Finally, it is critical that we understand what information is actually used in decision-making and how, and collect data reflecting these metrics.

## Conclusion

There is clear need for credible, timely research on the impact of policy and legal change for evidence-based policy-making in the innovation space. Greater engagement between patent offices, policy-makers and researchers is needed to ensure access to and collection of the most meaningful data. Equally, greater transparency in the field overall is critical. Researchers need to make readily available the datasets, analytical methodologies and assumptions underlying their findings. This would permit verification by peers, deeper understanding and follow-on research, thus promoting the efficiency, reliability and credibility of this field.