

Roundtable Report

Creating ATLAS

New Ideas for Leveraging
Publication and Patent Data
to Understand S&T Activity

Roundtable Participants

Shruti Agarwal

Catherine Aiken*

Melissa Flagg

Ronnie Kinoshita*

Christopher S. Marcum

Matt Marx

Shlomi Pasternak

Avital Percher

Katherine Quinn

Dave Rench McCauley

Caleb Smith

Patrick Thomas

Michael Toney

Andrew A. Toole

Allison Yezril

Grace Yuan

* Roundtable report author

On May 28, 2026, CSET hosted a roundtable entitled “Creating ATLAS: New Ideas for Leveraging Publication and Patent Data to Understand S&T Activity.” The roundtable brought together 16 experts from government, industry, academia, and nonprofit organizations for a dedicated discussion of two motivating questions:

1. What are new ways we can connect publication and patent data to derive useful insights into science and technology (S&T) activity?
2. How can we characterize these connections, visually or in natural language?

The roundtable was part of CSET’s recently launched [Project ATLAS](#). Project ATLAS is led by CSET’s [data team](#) and supported by [Google.org](#). It aims to increase the utility and interpretability of scientific research data, evaluate and implement LLM-based methods for extracting research metadata, organize and describe the global scientific landscape, and identify emerging areas of activity. As part of the project, CSET is building use-inspired open research infrastructure, including the enhancement of CSET’s [Map of Science](#), and developing new [open datasets](#) and [tools](#) to enable better exploration and identification of emerging areas of science and technology, increase access to scientific data and insights, and foster knowledge sharing.

The primary goal of the roundtable was to surface and discuss new approaches to analyze, connect, and visualize scientific publication and patent data to deliver insights and inform S&T decision-making. Additional goals included building relationships, fostering information sharing, and collaboratively brainstorming concrete ideas for research and tool development.

Discussion 1: Connecting Publications and Patents

The first motivating question asked participants to discuss new ways we can connect publication and patent data to derive useful insights into science and technology (S&T) activity. The discussion surfaced several methodological and conceptual challenges when relying on publication and patent data to discern genuine S&T signals and insights. Such challenges range from resolving the varied terminology of scientific publications and legal documents to confronting the strategic human behavior behind research and patent citation practices.

Participants emphasized that identifying meaningful connections requires moving beyond singular links like text matches or citations toward diffusion pathways and “knowledge chains” that bring together multiple links to reveal the paths connecting research and patents and communicate the steps in and the content of those paths. Participants also argued for looking at the absence of connections as a way to identify

gaps, [structural holes](#), or potential missed opportunities in the innovation landscape and analyze why they exist. Analyzing these gaps may reveal areas where research has stalled, is strategically hidden, or lacks translation, which can help decision-makers identify high-impact areas for intervention and guide future investment and policy decisions.

Additional themes from this discussion included:

- **Begin by defining the purpose, and stay open.** The utility of any dataset or tool is dictated by its end-use case. Participants stressed that before engineering connections between datasets, it is best to identify the target audience—whether policymakers needing actionable intelligence or researchers seeking new collaboration—and the specific information gaps we aim to address. Participants also flagged the value in applying [FAIR](#) data principles into open tool design to enable the broadest set of users who may be able to make use of the tool, including for beneficial purposes not envisioned by developers.
- **Human validation still matters.** While more data, compute, and AI models allow us to identify and analyze connections more quickly and at a bigger scale than before, there is a continued need for human oversight in data interpretation and responsible AI use. This warrants a focus on equipping humans to evaluate and validate data and downstream claims—both in terms of developers of data and tools and the users leveraging them to make informed decisions.

Discussion 2: Visualizing and Communicating Connections

The second motivating question asked participants to share ideas for how we can characterize observed connections, visually or in natural language. Given the complexity of resulting connections and pathways, how they are visualized and narratively described is important for utility and interpretability. Participants again focused on transparency, user validation, and contextualization. They recommended that tools move beyond simple visualizations to include contextual information and “claim ladders” that allow users to audit the data and logic behind an insight—for example, a narrative overview of the citations connecting one set of publications to another and the substance of those connecting publications, with direct access to the sources summarized and methods used to create that overview (and historical data when those overviews change as new information comes in).

Other recommendations included visually highlighting “blind spots” and gaps and enabling comparative analysis through interfaces and stories. The conversation

concluded with a focus on usability, suggesting that to bridge the gap between data and policy-level decision-making, tools could employ AI-assisted, persona-driven interfaces and provide curated outputs, including case-specific data narratives, rather than relying on raw data or complex maps that may be inaccessible to many users.

Additional themes from this discussion include:

- **Build auditability into visualizations.** To move from data to defensible claims or policy, participants advocated visual and narrative outputs that make the interpretation process explicit and include continuous checks and updates as new information is added. This allows users to trace the path from data to inferred relationships to policy implications, ensuring that claims are transparent and easy to verify.
- **Know the audience.** A central challenge is ensuring that visualizations and tools are useful for broad audiences of researchers, policymakers and decision-makers. Suggestions to address this included implementing AI-driven “prompt aides” to help users navigate decision spaces and publishing curated outputs for specific gaps or use cases.

Key Takeaways

Several key points surfaced over the course of the roundtable discussions. These high-level takeaways suggest new paths and priorities for ongoing and future work to make publication and patent data more useful and timely for S&T analysis and decision-making.

Start with the desired insights. This tried-and-true approach to research and tool development reminds us to start with the question or need that we aim to address. Publication and patent data can tell us a lot but comes with many known limitations. While it may be useful to debate the merits and use of such data conceptually, it is more useful to first define the information gap we need to fill and the decisions that need to be made—in this case, specific to S&T policy—and evaluate how the data can help us fill that gap or make that decision.

S&T use cases that came up in the roundtable discussion included:

- Surface new links between research field or patent categories
- Describe the current S&T landscape, or specific areas of interest, and its evolution over time
- Generate new ideas or hypotheses to inspire breakthroughs
- Identify hubs or centers of excellence for engagement and collaboration
- Assess the impact of funding and interventions
- Inform regional capacity building and economic development
- Monitor and leverage national S&T advantages

Make it Useful. Users from researchers to policymakers want information that can be actionable in minutes. They want to engage with narratives and clear signals that they can trust. Datasets, tools, and interfaces need to provide clear, auditable information and be built for purpose.

Bibliometric data outputs or tool features that were suggested in the roundtable included:

- Curated narratives like briefs on national or regional S&T capabilities
- Research-to-impact timelines
- Signal alert systems
- Regional maps and capacity trackers
- Blind spot tracker, or “opportunity” spotter
- Research-to-patent pathway finder
- AI-enabled chat with user personas to provide decision space guidance

Make validation and auditability a priority. Participants encouraged a focus on mapping the connections, pathways, and chains linking publications and patents, highlighting that recent methods allow us to trace such paths at scale and make them visible to users. By making the connections transparent and visible to users, we can enable them to assess the assumptions built into the data and methods and to confirm their understanding and confidence in resulting claims. This is critical for trust, as more information is available than ever, yet establishing that trust grows more difficult. If our goal is to enable users to make defensible claims using data, we need to allow users to understand how the claims came to be and calibrate their trust in them over time.

Bring time into play. Related to validating and establishing trust, participants encouraged a focus on building in real-time updates of not just underlying data, but of the connections, graphs, and claims that result from them. Building in the timing component involves timestamping and clear version history for the data, but also for the downstream predictions, recommendations, and narrative descriptions of observed trends and connections. This will allow users to update conclusions as new information becomes available, driving more accuracy and confidence in data for S&T analysis.

Leverage datasets in combination. While individual publication and patent datasets continue to provide insights, changing real-world trends and advanced methods should push us to lean into what they tell us in combination. This goes beyond publication and patent data, to include other sources of information relevant to S&T development like social media, grant and funding information, and company filings. This suggests a focus on connecting data sources and analyzing the linkages and pathways, or absence thereof, between them.

Recognize changing trends and special cases. There is no "one-size-fits-all" signal for novel S&T developments. We must be cognizant of "special" cases where science or technology may not be developing along historically observed pathways, or disrupting the pathways that historically defined connections. AI and quantum were discussed as two cases where the expected, observable flow of knowledge from research to patenting may not apply. Also, timelines may be technology or field specific and need to be accounted for.

What Is Next?

The roundtable offered ideas for Project ATLAS, as well as recommendations for the broader S&T research and tool development community, which includes government analysts and service providers, as well as industry practitioners and leaders, and academic metascience, innovation, and technology researchers. We hope the themes surfaced in our discussion help others in this space develop use-inspired data outputs, apply new methods to longstanding questions, and advance collective understanding of S&T landscape and dynamics. We also hope to spur continued discussion. As we work on these topics, we welcome ideas, feedback, and collaborations.

For the remainder of this year, the Project ATLAS team is focusing on:

- Getting feedback on our newly launched [Map of Patents](#) on our Emerging Technology Observatory.
- Sharing the new topic areas added to our [Research Almanac](#), covering more AIxBio, cybersecurity, and chip research.

- Improving bibliometric metadata, including venues, funders, and subjects.
- Building “research diffusion pathways” using links within and across the publication and patent clusters and using LLMs to summarize the pathways.
- Identifying research clusters with high recent research activity.
- Assessing and describing research cluster scope, collaboration, and interdisciplinarity.
- Research funder portfolio analysis and impact assessment.
- Designing and prototyping new interfaces, including overlays for emerging areas and real-time monitoring.

If you are doing related work and interested in exploring collaborative opportunities with CSET, please contact us via email at cset@georgetown.edu.

Acknowledgments

The authors thank the following roundtable participants for their time and contributions.

Shruti Agarwal, Center for Security and Emerging Technology

* Catherine Aiken, Center for Security and Emerging Technology

Melissa Flagg, The Andrew W. Marshall Foundation; Syntheos

* Ronnie Kinoshita, Center for Security and Emerging Technology

Christopher S. Marcum, Federation of American Scientists

Matt Marx, Cornell University; National Bureau of Economic Research's Innovation Information Initiative (i3)

Shlomi Pasternak, Google Research

Avital Percher, National Science Policy Network

Katherine Quinn, Center for Security and Emerging Technology

Dave Rench McCauley, Gridium

Caleb Smith, University of Michigan; Syntheos; SciTech Strategies

Patrick Thomas, 1790 Analytics

Michael Toney, Google Research

Andrew A. Toole, Georgia Institute of Technology

Allison Yezril, Decision Space LLC; retired intelligence community expert

Grace Yuan, National Science Foundation

**Roundtable report author.*



© 2026 by the Center for Security and Emerging Technology. This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc/4.0/>.

Document Identifier: doi: 10.51593/20260003