

Making Grey Literature Discoverable and Impactful on JSTOR Through Comprehensive Search and Rich Metadata

Robert E. Sedgwick and Rachel Ross

Robert Sedgwick: Content Development & Licensing (JSTOR), ITHAKA, New York, NY, USA;

Rachel Ross: Content Management (JSTOR), ITHAKA, New York, NY, USA

Robert Sedgwick

(Corresponding Author)

Senior Editor, Grey Literature (JSTOR)

ITHAKA

101 Greenwich Street, 18th floor

New York, NY 10006

Tel: 212-358-6431

Cell: 646-670-8570

Robert.Sedgwick@ithaka.org

Robert Sedgwick is Senior Editor, Grey Literature at JSTOR; Rachel Ross is a Metadata Librarian at JSTOR

Making Grey Literature Discoverable and Impactful on JSTOR Through Comprehensive Search and Rich Metadata

ABSTRACT. Think tank grey literature can be used for teaching and research in policy-oriented courses, but because much of this content is scattered throughout the far reaches of cyberspace, students and many faculty may not know how to go about finding it. At the same time, think tanks need to find new and creative ways to push their content out and capture new audiences. This article will assess the value of aggregating grey literature on JSTOR for both researchers and think tanks, briefly touch upon the advantages of comprehensive searching on our platform, and then take a detailed look at how we make this content discoverable through the application of rich metadata.

Keywords: grey literature, information discovery, metadata, research institutes, user experience

Introduction

Think tank grey literatureⁱ (encompassing working papers, policy briefs, occasional papers, monographs, etc.) has long been considered a valuable resource for teaching public policy and policy-oriented classes and supporting research at colleges and universities. The accessibility of these reports and their proclivity to cluster around the cutting edge of current research topics make them popular among faculty and students alike. However, it is often difficult to discover this content using conventional Google searches, and digging through siloed repositories on individual websites can be both cumbersome and time consuming. As a result, many of the best research reports from

reputable think tanks slip through the digital cracks and go largely unnoticed.

JSTOR has attempted to redress these issues by centralizing a curated collection of think tank research reports on a single platform, making this content freely available to all JSTOR users, and enhancing its discoverability through comprehensive searching and the application of rich metadata. This not only benefits university-based researchers, it also serves the interests of the think tanks themselves as they look for ways to stay relevant in today's fast-paced and technology-driven media environment and reach new audiences beyond the policymaking community.

Breaking Through the Clutter and Staying Relevant

There are currently more than 8,000 think tanksⁱⁱ internationally according to the *2019 Global Go To Think Tank Index Report*, with the lion's share of these institutes based in the United States (India, China and the United Kingdom come in second, third and fourth respectively).ⁱⁱⁱ Since the early 1970s, the number of think tanks has grown exponentially as world events and the issues shaping them have become more complex. There was also (in the 1970s and 1980s) a shift away from the traditional, research-based institutes, which exercised a certain degree of ideological detachment, toward a newer breed of think tank that was highly partisan and sometimes attached to a political party.^{iv} In the U.S., these "advocacy tanks",^v as they are often referred to, became unabashedly activist in subsequent decades as the country became more politically polarized.^{vi} For many years, think tanks were considered reliable sources of expertise by policymakers and journalists who actively read their content and consulted their fellows on the issues. But as these institutes entered the new millennium, their influence began to wane as they struggled to adapt to the new online environment. Although the Internet provided them with a larger platform to reach their audiences,

they now faced growing competition from alternative electronic sources, including highly partisan news aggregators and blogs, social media platforms, not to mention the 24-hour cable news stations--all vying for the attention of policymakers who found themselves suddenly overwhelmed by a torrent of available information.^{vii} In short, think tanks were finding it increasingly difficult to break through all the clutter and get their ideas heard.

So how do think tanks, and particularly the more traditional research-based institutes, stay relevant in a politically contentious and data-rich environment, and become technologically innovative enough to broaden their reach? One strategy is to look beyond the policymaking world to other audiences that have an appetite for the kind of content that think tanks produce. Colleges and universities throughout the world, with millions of faculty and students, appear to be the ideal ancillary market for this grey literature. Unfortunately, most think tanks are not-for-profit organizations with limited budgets and do not have the resources to allocate toward developing and reaching new audiences. It therefore makes sense that many of them would outsource this task to content aggregators with large and recognizable platforms.^{viii}

Open Research Reports on JSTOR

For years now, librarians have been inundated with queries about think tank research reports. Many of these requests come from students who had heard or read about a recently published report from a think tank cited in the news but did not know how to go about finding it. Although much of this grey literature has been online since the early days of the Internet, it was not easily discoverable and students struggled to discern the best and most credible materials amidst a growing corpus of questionable sources. Moreover, even if you knew where to find legitimate reports to support your research, searching across different sites was not ideal, with each think tank having its

own unique way of surfacing content.

At JSTOR, we speculated that our users would find value in this content if it was curated and made open and easily discoverable on our platform alongside relevant journal articles and book chapters. However, we needed to test this hypothesis. Before making think tank research reports widely available on JSTOR, we first assessed their value, based on usage and end user feedback, by limiting their access to institutions that licensed two of our thematic collections: *Security Studies* and *Sustainability*. Over a three-year period, we found that usage for these reports was at least on par with the journal articles included in these collections. We also heard from many faculty during this time about the convenience and benefits of having a curated collection of think tank reports available in one place.

Having thus confirmed this value within the academic community, and with the support of the think tanks we work with, we converted all of our grey literature to open content, making it freely available to all JSTOR users as part of our mission to expand access to teaching and learning. Within the first two months of launching our open reports initiative in February 2020, we saw the usage of this content increase by more than 800 percent. On average, research reports have been generating more than 500,000 item requests (e.g. article views and downloads) per month on JSTOR.

The Value Added

As of October 2020, there are just over 25,000 research reports on JSTOR from more than 100 think tanks. Existing content is updated regularly while new institutes are selected based on faculty recommendations and in-house research.^{ix} Although we initially limited the scope of these reports to support our *Security Studies* and *Sustainability* collections, we recently licensed and released more than 600 reports covering health and health care topics including COVID-19 in response to the current

pandemic and are looking to add reports in other disciplines as well.

Archiving a curated collection of grey literature on a single platform and making it discoverable benefits students and faculty in myriad ways. First, it helps ensure that students are only accessing the most legitimate and relevant reports to support their research. As many students tend to begin their research with Google or Wikipedia, faculty are only too happy to send their undergrads to a trusted source, such as JSTOR, to look for content. And faculty find value in having a centralized collection of think tank reports available to support their own research needs as well. “It’s not so much about the journals because I know which ones I like to use and where to find them,” explained Professor Carlo Masala of Bundeswehr University Munich recently when running searches within our *Security Studies* collection. “It’s more about the grey literature. Collecting this content in one place and making it discoverable is research gold because most people don’t know where to go to get this kind of content.”

Secondly, the international and ideological diversity of these reports provides students with real-world perspectives on many of today’s most salient issues and debates, including climate change, terrorism, health care, geopolitics, and cybersecurity. To that end, we have made a concerted effort to include reports published by think tanks situated in parts of Africa, Asia, and Latin America to give those institutes, and those regions, a voice in these policy discussions. In fact, more than half of all our contributing think tanks are currently based in countries outside the United States. Exposure to a multiplicity of viewpoints from different cultures and from various ends of the political spectrum helps to foster both debate and empathy within the classroom, thereby providing students with a broader educational experience wherever these reports are used.

The currency of this content is another benefit. Think tank research reports are

generally more up-to-date on policy issues compared with peer-reviewed journal articles and books. Some think tanks, such as the Center for Strategic and International Studies, for example, publish up to a dozen, or more, new reports each month. This makes this content the perfect supplement to the more traditional in-classroom research materials used by faculty.

Research reports are also more accessible to a broader audience as they remain relatively free of the kind of academic jargon that characterizes many of the journal articles assigned to students. Think tanks are adept at distilling dense academic research into concise publications that are palatable to government decision makers since the overriding objective of this output is to potentially impact policy.^x

Yet another advantage is archival preservation. Many think tanks, especially those in the developing world, are facing serious threats. In Turkey, for instance, the government shuttered several research institutes after the failed coup in 2016, and the dearth of funding in many countries is another problem. Think tanks in Africa are disappearing at an alarming rate, with many institutes from that continent not expected to survive the next five-to-seven years.^{xi} As a result, much of the content from defunct think tanks is sadly lost to the scholarly community forever. Having these publications perpetually archived somewhere guarantees their preservation. For example, although several think tanks that signed with us have since closed down, the back runs of their content will always be freely accessible on JSTOR.

Search Discovery and Metadata

The JSTOR search experience facilitates research by connecting think tank grey literature with other relevant content types such as journal articles and book chapters within a given body of content. A user, for example, may be searching for a specific think tank report on JSTOR only to be pleasantly surprised to discover in the search

results journal articles and/or book chapters that also support that user's research needs.

“The true value of this resource [JSTOR] is that it provides you with a whole set of content adjacent to the content you are looking for, and it's often content you didn't know about,” said Professor Arjun Chowdhury of the University of British Columbia.

Discoverability of research reports within the thousands of journals and ebooks on JSTOR requires the capture of rich metadata. In addition to OCRing the content, we send the research reports to an outside vendor to create the XML metadata files. The vendor uses our in-house metadata specification, based on the NLM's Book DTD 3.0, and works closely with JSTOR's Metadata Librarians to ensure the correct capture of elements.

Metadata for the research reports is captured at both the report level and the “chapter” level. Report level metadata elements include the research report title, series title, contributor and contributor role information, copyright statements and years, publication dates, publisher information, and (when printed in the source) LC subject headings and ISSN and/or ISBN information [Figure 1 near here].

At the chapter level, we capture titles, contributors, pagination (i.e., first and last page numbers), extracts, and content type information for each smaller part of the report.

These elements are leveraged by JSTOR in many different ways; they are displayed in the user interface, used to generate citations for users, added to the JSTOR search index to aid in more precise discovery^{xii}, and are sent to third-party discovery partners to populate their discovery layers.

Moreover, metadata properly parsed into standardized elements facilitates linked data throughout the Web in general, making the research reports not only easier to find on JSTOR but also within larger search engines. A deeper dive into some of these

elements shows the detailed capture of the metadata and how it is used to drive search and discovery.

For high-impact metadata elements such as title and contributor information, we capture metadata for the report as a whole (i.e., the information on the cover or title page) and for the organizational units of the report, typically called chapters or book parts (i.e., chapter titles and chapter authors). Separating out and capturing metadata for these smaller units of the report allows for more unique title and author information to be added to the search index, resulting in more focused searches, better discoverability of content, and the surfacing of more targeted sections of the report. Parsing the content into more granular elements in the XML, such as title, subtitle, and trans-title, and for contributors, given names and surnames, is also useful in discovery, both in the JSTOR search index and with creating stronger inbound linking. Additionally, we capture a publication date for every report. This metadata point is incredibly important in grey literature, and in turn lets the user scope and facet their search/search results by date.

Search is not the only feature that benefits from the robust metadata capture, the user experience is also enhanced because of it. When reports are broken into smaller chapters, the user has the ability to click through the report's full table of contents to help facilitate the user's navigation within the report. This functionality is not inherent in the reports themselves; it is a result of the metadata capture.

The JSTOR thesaurus is also a crucial part of the user experience, and is metadata driven. The thesaurus is internally designed and maintained by in-house Semantic Librarians and Subject Matter Experts, and organizes content by subject matter. Subjects, or topics, are assigned to reports based on how many times a particular term is present in the source. The topics are clickable and provide users with the option to explore other JSTOR content that is similarly categorized, or search within a topic to

refine their results even further. Being able to click through reports that have been grouped by similar topics helps the user find more content relevant to their search, benefiting the user experience and increasing the discoverability of other reports.

Conclusion

We saw robust usage of open access books and journals when we added them to JSTOR, and are experiencing similar success with open research reports. As researchers look for more resources online, JSTOR will continue to explore new content types and partner with the institutions that produce them.

Think tanks are currently facing two challenges: trying to stay relevant and influential in a vast digital world dominated by social media giants; and devising new ways to disseminate their content and gain traction with larger and newer audiences. Platforms such as JSTOR can help with the second challenge in particular by introducing this content to undergraduate students who might not be aware of it, or might not know how to find it.

For faculty and students, we are providing a rich corpus of open content that is both ideologically and internationally diverse in addition to bridging the gap between the academy and the policymaking world. Think tank research reports are not by any means a substitute for peer-reviewed journal articles or book chapters, but they can be used to supplement these materials in policy-driven courses and are generally valued for their currency, accessibility, and their ability to deliver concise analysis on many of today's most hotly debated topics.

Figure 1. *Research report cover image and snippet of corresponding report level XML markup*

Notes

ⁱ I use the terms “grey literature” and “research reports” interchangeably throughout this article.

ⁱⁱ James McGann defines think tanks as “*organizations that generate policy-oriented research, analysis, and advice on domestic and international issues that enable policymakers and the public to make informed decisions about public policy issues.*” James McGann, “Think Tanks and the Transnationalization of Foreign Policy,” *Foreign Policy Agenda* 7, no. 3 (2002), 1-2. Some well-known think tanks would include for example: Brookings Institution, Chatham House, CATO, Council on Foreign Relations, Institut Français des Relations Internationales, and RAND Corporation. A number of think tanks produce peer-reviewed journals (e.g. the *RAND Journal of Economics* and the *CATO Journal*), many present papers at academic conferences and they generate research in a variety of disciplines including economic development, education, foreign policy & international relations, health care, political economy and socio-political issues.

ⁱⁱⁱ James McGann, “2019 Global Go To Think Tank Index Report” (Think Tanks & Civil Societies Program, The Lauder Institute, The University of Pennsylvania, Philadelphia, 2020), 43.

^{iv} Andrew Rich, “Think tanks and policy analysis,” in *Policy analysis in the United States*, ed. John A. Hird (Bristol, UK: Bristol University Press, 2018), 284-85.

^v Brookings Institution, Carnegie Endowment for International Peace, and RAND Corporation are examples of traditional, research-based think tanks while institutes such as CATO, the Center for American Progress, and The Heritage Foundation could all be considered advocacy tanks.

^{vi} Tevi Troy, “Devaluing the Think Tank,” *National Affairs*, no. 10 (2012).

^{vii} James McGann, "For think tanks, it's either innovate or die," *The Washington Post*, October 6, 2015.

^{viii} Other platforms, besides JSTOR, that host think tank grey literature include, for example, Columbia International Affairs Online (CIAO), EBSCO and ProQuest.

^{ix} Subject matter experts at JSTOR put together a target list of potential think tanks for a given collection; those think tanks are then vetted both in-house and with faculty we engage with by phone, electronically or at conferences. We interviewed, for example, more than 200 faculty for the *JSTOR Security Studies* collection for feedback and content recommendations. We also rely on the guidance of advisory groups, which include faculty and working professionals in a given field, and assigned to a specific collection.

^x Rohinton Medhora and John de Boer, "Think tanks exist to influence: A conversation with Rohinton Medhora," *International Journal*, Vol. 70, No. 4 (2015), 2.

^{xi} James McGann, Landry Signé, and Monde Muyangwa, "The crisis of African think tanks: Challenges and solutions", (Brookings Institution, Washington, DC, 2017).

^{xiii} Laura Waugh, Hannah Tarver, Mark Edward Phillips, and Daniel Gelaw Alemneh, "Comparison of Full-text Versus Metadata Searching in an Institutional Repository: Case Study of the UNT Scholarly Works," University of North Texas Libraries, UNT Digital Library, last modified February 8, 2015, accessed August 30, 2020, <https://digital.library.unt.edu/ark:/67531/metadc725823/>.