Validating a Pipeline to Create a Comparable Corpus of Government-Issued Travel Advisories from the Internet Archives

Laura Braun, Christian Oswald
University of the Bundeswehr Munich
Neubiberg, Bavaria, Germany
laura.braun@unibw.de, christian.oswald@unibw.de

Abstract

Government-issued travel advisories are used by citizens to get information about destination countries for tourism and other purposes such as temporary work stays or permanent relocation plans. However, qualitative evidence suggests that travel advisories may be influenced by considerations beyond current security situations. Systematic and rigorous quantitative analyses of advisories are scarce because relevant corpus data are not readily available and official government websites often provide practical obstacles. We validate a pipeline to generate a time-series cross-sectional dataset of government-issued travel advisories for three English-speaking issuing countries based on the Internet Archive's Wayback Machine. Using official government data sources that are prohibited to be scraped and used for research, we illustrate that our approach provides (near-)complete coverage. The resulting corpus and code are intended to support downstream research on comparative risk communication, international relations, and text analysis using natural language processing methods.

Keywords: travel advisories, corpus comparison, web archives, corpus construction, time-series cross-sectional data

1. Introduction

Travel advisories are designed to protect citizens from potential hazards and inform them about risks abroad. They also affect economic and political developments in the target countries of these advisories. Qualitative evidence suggests that travel advisories may also be in place for political or strategic reasons (Sharpley et al., 1996; Babey, 2019; Chu et al., 2021). However, systematic quantitative evidence for any such links is absent. Obstacles in obtaining data on travel advisories are numerous. Travel advisories are usually posted on government-administered websites. Website structures and URLs, as well as the structure of the content itself, can change over time. The content is regularly updated and replaced over time. Many of these websites also explicitly forbid the use of web scraping to obtain data and do not provide API access. Time-series cross-sectional data on government-issued travel advisories would be a valuable data source for research areas as diverse as tourism, communication, social sciences, or humanities and can be used for qualitative, quantitative, and text analysis approaches. We evaluate a transparent and reproducible pipeline to generate such a dataset for several issuing countries based on the Internet Archive's Wayback Machine (IAWM) using official government sources.

2. Background & Summary

Travel advisories and warnings show how governments communicate risks abroad. The web has been the main platform to inform the public about risks since the late 1990s (Löwenheim, 2007). Previous work links advisories to, for example, tourist behavior and policy communication in international relations (Sharpley et al., 1996; Murphy et al., 2007; Babey, 2019; Chu et al., 2021). However, these are mainly case studies and there is no global longitudinal dataset that spans various issuers, languages, and destinations. Previous research addressed this gap and designed a data collection pipeline to create time series by issuer and target (Braun and Oswald, 2025). We validate these efforts and estimate the validity of this approach and the resulting data to ensure that the travel advisory data can be reliably used in various research areas. With a multi-issuer corpus in place, research questions that were previously out of reach become possible to answer. The data allow to test, for example, convergence and divergence in risk communications between issuers, to track risk communication behavior over time, or whether the advisory represents the political and economical relationship between the issuer and the destination rather than the actual security situation on the ground.

The Internet Archive's Wayback Machine allows us to analyze web content over time as it changes (Weikum et al., 2011). URL-timestamp lists and stable snapshots allow us to reconstruct advisories that were already overwritten a long time ago. At the same time, web archives are neither neutral nor pure; they simply do not provide a fully systematic mirror of historical online content. Rather, they are socio-technical systems whose coverage results from a mix of broad crawls, institutional collections, hyperlink ecology, and individual archiving requests (Ben-David and Amram, 2018). English is the lan-

Proceedings of the 19th Workshop on Building and Using Comparable Corpora (BUCC) @ LREC 2026, pages 96–107
11 May 2026.

guage with the highest visibility, with other European languages and Japanese following (AlNoamany et al., 2014). These patterns create an unbalanced snapshot density between issuers. In addition, the timing of snapshots varies and it can happen that an update may be captured days, weeks, or even months later. Delays between the actual content update on the web page and the time of the snapshot can hide interim updates. Access can also be affected by robot policies, site blocking, and even the location of the IAWM's server (Ben-David and Amram, 2018). All these aspects lie in the nature of the Internet Archive, but might influence the completeness of the collected data. However, the IAWM presents the best approach for obtaining cross-sectional time-series data of government-issued travel advisories in a reproducible and transparent manner in the absence of viable alternatives to obtain travel advisories at scale.

Table 1: Number of snapshots of index page with HTTP 200 responses

Country	Archive start	# IAWM 200
Australia	1997	1675
Canada (fr)	1997	1769
Canada (en)	1997	4902
China	2005	507
France	2000	2628
Germany	1998	1905
Hong Kong (ch)	2009	33
Hong Kong (en)	2009	256
India[1]	2016	241
Indonesia	2017	46
Japan	2003	2992
Mexico[2]	2003	193
Russia[3]	2016	109
United Kingdom	1997	10838
United States	1996	13342

3. Data and Corpus Construction

Our ultimate goal is to create a multilingual cross-sectional time-series corpus of government travel advisories that covers issuers on a globally representative scale. We build on a previously proposed TRAVELWARN-Crawler pipeline that can be adapted to additional issuers and languages (Braun and Oswald, 2025). We rely solely on the IAWM for our corpus construction. The IAWM offers dense coverage from the mid-2000s onward and an open CDX API for URL–timestamp enumeration, which makes it the most practical backbone for standardized retrieval over issuers and time (Murphy et al., 2007; Arora et al., 2016).

Country-specific URLs are gathered by manually identifying a list of seed URLs, which are the index pages that link to country pages (see the Australia index page example in Figures 8a and 8b). "Index page" refers to the main landing page of the travel advisory section, which then redirects to the country-specific warnings via links and, for some issuers such as the US and the UK, even presents a tabular overview with the latest update date for each country. We crawl these seeds for each issuer era and parse all links that point to country-level advisories into a database. An overview of manually identified URLs for each era and country of interest is provided in Table 3 in the Appendix. The pipeline then uses the country URLs with era boundaries. Using the CDX API for URL–timestamp enumeration, the crawler iterates over all snapshots for each country URL and parses content into the database whenever there is an unseen update date. This approach preserves overwritten content and supports longitudinal analysis.

Issuers differ in how they publish and present advisories on their index pages. Some expose the country link together with the latest update date, for example, US, Canada, France, and most of the times Australia (except around the year 2000). Others, such as the UK and Germany, do not. The Internet Archive tends to provide more snapshots of index pages than of country-specific URLs. For illustration, `https://www.smartraveller.gov.au/destinations/africa/sudan` was saved 98 times between December 3, 2019 and June 13, 2025, while `https://www.smartraveller.gov.au/destinations` was saved 545 times between December 9, 2019 and September 22, 2025. Using index page information allows us to generate validation data, although full recall is still not guaranteed.

We extract a minimal schema from each snapshot that supports replication and downstream text analysis. We store an index collection with at least the country name and country URL for each issuer and, where available, update dates or travel warning level information. We also store a text collection with country, source URL, warning date, advisory text, and where available additional metadata such as level and risk tags (Braun and Oswald, 2025). The design supports later comparisons of frequency and content between issuers.

Table 1 summarizes the total captures of index pages by all issuers of interest. The table highlights a strong imbalance in snapshot counts among issuers. The US index pages have by far the most snapshots. The UK also has many snapshots. Canada, Australia, France, and Germany are in an intermediate range. The French version of the Canadian pages has less than half the amount of snapshots as the English version. China has significantly fewer snapshots (Thelwall and Vaughan, 2004). These observations are in line with earlier

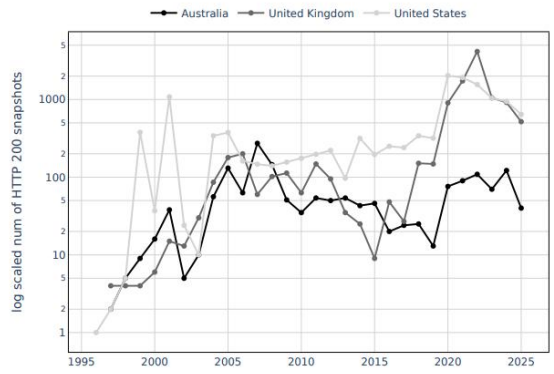


Figure 1: Yearly log-scaled number of HTTP 200 snapshots of travel advisory index pages

findings on the dominance of English language content and a Eurocentric and Western representation in web archives. We take this into account in the evaluation of the limitations of the IAWM as our data source. Table 3 in the Appendix provides a more detailed overview of the number of snapshots for each era, highlighting that the late 1990s and early 2000s have comparatively fewer captures.

Additionally, annually aggregated information on the number of HTTP 200 snapshots for the United States, UK, and Australia is plotted in Figure 1. We The values are shown on a logarithmic scale to account for the large differences in snapshot intensity between issuers and over time. Archiving for the US started in 1996 and soon reached a high snapshot density, reflecting an already high crawl intensity of its travel advisory index pages from the early 2000s onward. The UK and Australia first appeared in 1997, with fewer than ten snapshots in that year. The coverage for both subsequently increases, especially in the early 2000s, where all three converge toward roughly one hundred snapshots per year. All issuers show pronounced spikes in the number of snapshots during years affected by the Covid-19 pandemic, which may reflect general increased traffic to the warning pages. The number of snapshots increases to more than 4000 in 2022 for the UK, resulting in even several captures per day.

Note that the tables and line plot allow us to analyze the archiving behavior on the index pages only. We use country-specific pages for the data validation later in the paper, which in most cases have fewer snapshots than the corresponding index page. Analyzing the index page provides primarily a comparative estimate of how well a given issuing country is represented in the Internet Archive overall.

Table 2 provides a brief overview of the number of countries and territories for which each issuing country considered in this validation study

Table 2: Number of destination countries and territories covered by each issuer

Issuer	**# Countries & territories**
United Kingdom	230
Australia	221
United States	210

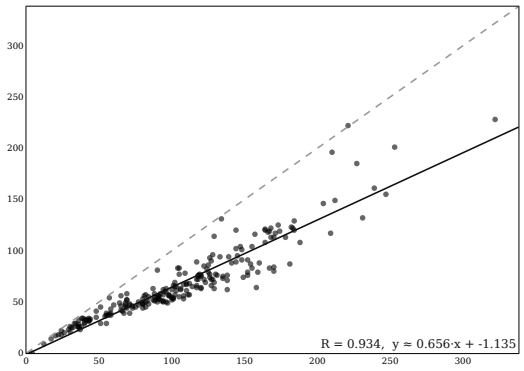
publishes travel advisories. The UK covers slightly more countries and territories than Australia and the US, as it includes a large number of island territories. The counts also include entities or states that no longer exist, e.g., Yugoslavia, as well as broader regional designations, such as “Africa (Central and Western)”, for which Australia issued a small number of advisories in the late 1990s. Given that the United Nations currently have 193 member and two non-member states, the pipeline seems to provide good coverage.

4. Validation

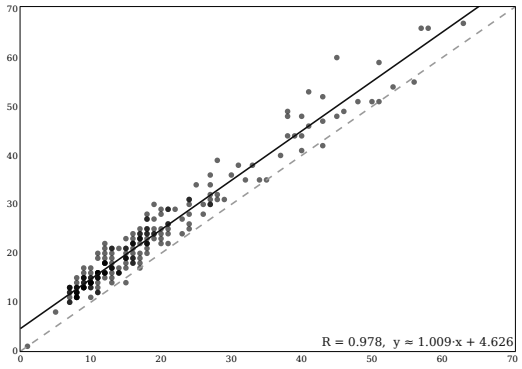
We validate the scraped data of three English-speaking issuing countries with two types of reference data. First, we derive validation targets from issuer index pages for Australia and the United States, which the Internet Archive captures more frequently than destination pages. Second, we use external sources where terms of use allow validation but not redistribution for the US and the UK.[4] We use items tagged as *Travel Advisory* from the Overseas Security Advisory Council (OSAC) catalog beginning in 2004 for the US. OSAC is a public-private partnership that seeks to help protect US interests overseas and is part of the Bureau of Diplomatic Security of the US Department of State. We use captures from the UK National Archives, similar to the IAWM, via a CDX endpoint and apply the same extraction logic as in Braun and Oswald (2025) for the UK. The National Archives also store the full timeline of published online advisories. Compared to the Internet Archive, their crawler nowadays runs systematically once every working day, but also has bigger gaps in earlier years. Note that the National Archives data can be used for validation purposes, but takedown and reclosure policies make scientific use and sharing of the data challenging.

We assess agreement and coverage. We fit an ordinary least squares regression model of our per country counts on the reference counts for matched years to measure agreement. Perfect agreement lies on the line $x = y$, and a larger R with a slope close to one indicates better alignment. We compare the annual number of advisories per destina-

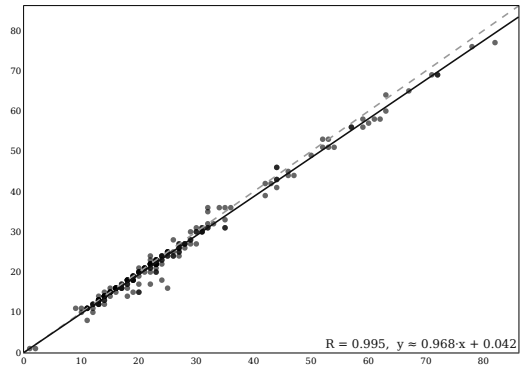
[4]We provide scripts to replicate the collection of the validation sources.



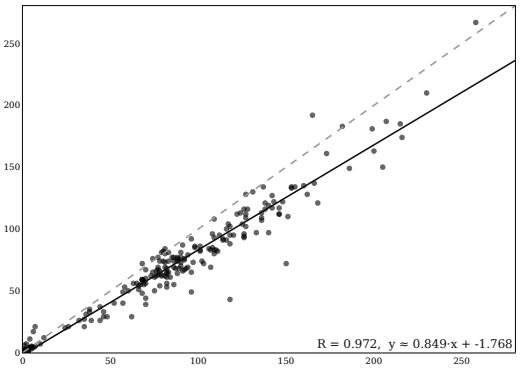
(a) UK vs. UK validation



(b) US vs. OSAC



(c) US vs. US index



(d) AU vs. AU index

Figure 2: Per country agreement between Internet Archive based counts of travel advisories and validation data. Each panel plots, for one selected country, the number of advisories in the reference data (horizontal axis) against the number recovered from the Internet Archive (vertical axis), together with the comparison line ($x = y$).

tion against the reference to measure coverage and summarize the distribution as violin plots with per year means. None of the reference data guaranties perfect recall. Our corpus can show overcoverage when it contains versions that are missing from the reference data. Validation is not affected by timing issues, as we base annual measurements only on the warning date rather than the snapshot date.

Figure 2a compares per-country advisory counts derived from the UK country pages in our Internet Archive corpus with counts based on the UK National Archives country pages. The regression line has a slope of about 0.66 and a negative intercept, which implies that we recover slightly more than two thirds of all advisories and that the shortfalls become larger for destinations with many updates. The annual coverage distribution in Figure 4a shows that this undercoverage fluctuates over time, with some earlier years showing wider negative tails, whereas later years, particularly 2022, cluster closer to the reference data. Figure 5 in the Appendix contrasts UK country pages as archived by the Internet Archive and by the UK

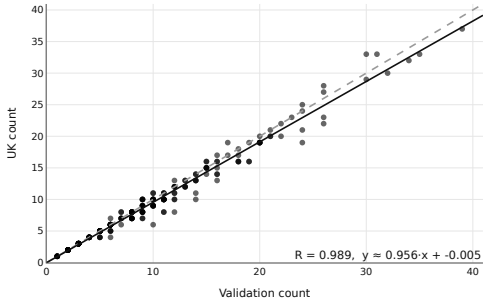


Figure 3: Agreement Internet Archive UK vs. The National Archives UK for the year 2022

National Archive only for the year 2022, a year in which the Internet Archive captures the index page almost daily, and with it seemingly also the country pages more often, as the almost perfect agreement indicates.

Although Figure 4a reveals a loss of updates for

years other than 2022, the correlation is high, and the coverage violins show no evidence of systematic exclusion of, e.g., a specific set of countries. Instead, missing observations are dispersed across destinations, so that relative differences between countries and overall temporal dynamics seem to be preserved even though absolute UK update counts are systematically underrepresented in our corpus.

The agreement between OSAC and our US corpus is strong with $R = 0.978$ and a slope close to one, as shown in Figure 2b. Notice the positive intercept of about $+4.6$, which indicates a small surplus on our side. Put differently, country level counts are systematically higher in our corpus, which in turn suggests that almost all destinations listed by OSAC are present in our text set and that we recover additional updates that do not appear in the OSAC catalog. The coverage plot in Figure 4b shows that this surplus is concentrated in recent years, especially between 2021 and 2023, where the OSAC catalog lists comparatively few entries. Screenshots of the OSAC catalog in the Appendix (Figure 6) highlight that filtering for 2023 only returns 91 entries compared to the year 2019 with more than 300 results. We do not have any insight into this gap, but it illustrates that collecting perfect data is challenging.

We also compare the US corpus to the US index derived target from the archive itself. Figure 2c shows an almost perfect fit with $R = 0.995$ and a slope of about 0.968. On average, about 3% of updates per country fall between captures of the index page. The coverage in Figure 4c is tightly centered near parity with a few expected outliers when a destination has only one advisory in the reference but none in the text set, which would translate to a coverage of -1. Given the large number of snapshots for US index and country pages and the moderate US update frequency compared to other issuers, we are confident to have a near-complete reconstruction for the US Department of State travel advisories.

Figure 2d shows high agreement between our scraped data and the Australian index target with $R = 0.972$ and a slope around 0.849. Undercoverage in our data grows for destinations with many updates, which fits the lower snapshot density of Australian country pages relative to index pages. The coverage violins in Figure 4d confirm wider negative tails in years with rapid update activity. The index design did not expose update dates around the year 2000 (see Figure 8a), which reduces the quality of the index-based target in that period. This explains a spike toward overcoverage in our text data when country pages carry advisory updates that the index does not surface.[5]

[5] Figure 2d does not reflect this design change, which may slightly lower R.

5. Discussion

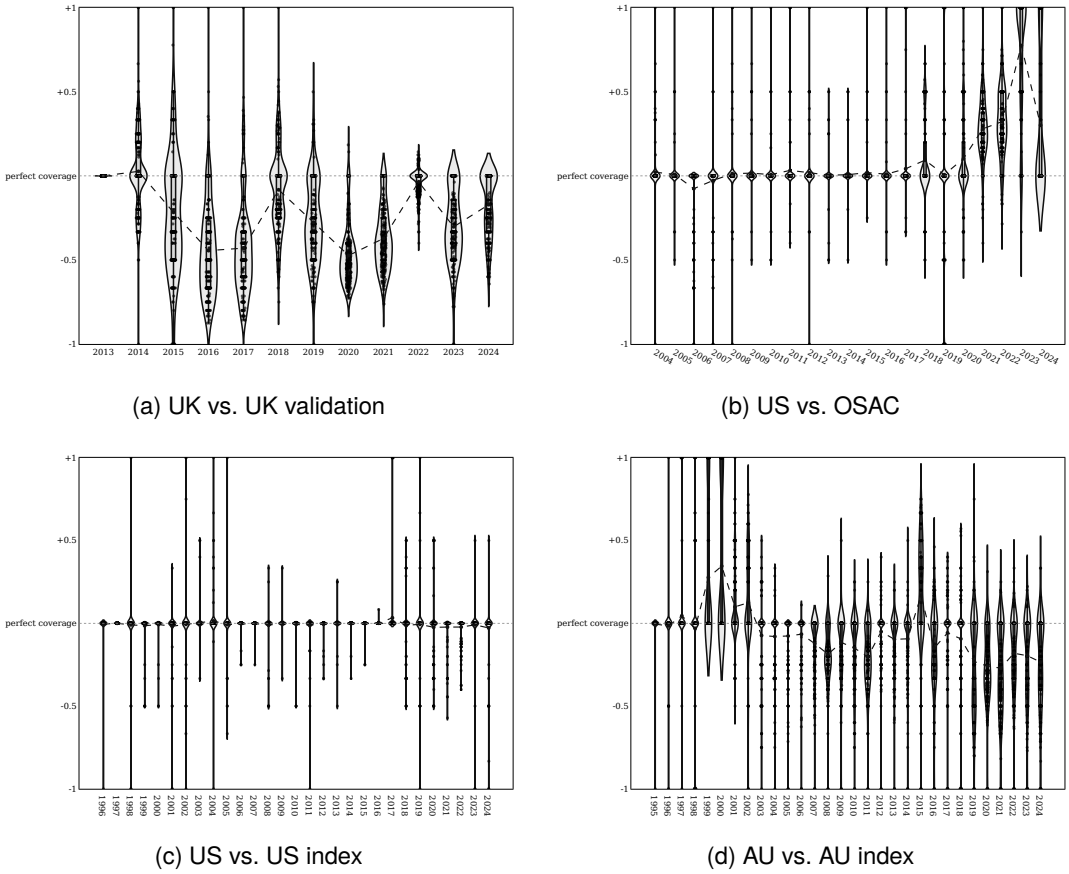
The validation results show that the IAWM, although not without limitations, serves as a reliable source for reconstructing advisory timelines at scale. We obtain a near complete reconstruction for US travel advisories. The agreement with OSAC and with the US index is high and coverage is close to complete. For the UK, comparing country-level counts to the UK National Archives reveals undercoverage in absolute terms, but also very high correlations and no evidence of a subset of destinations being dropped entirely. The UK and Australian results together suggest that the main weakness of our Internet Archive based approach lies in capturing issuers with very frequent updates, especially when snapshots are infrequent. Additionally, not only the total count of updates by country, but also previous work indicates that the US updates less frequently with more major changes per update than the UK and Australia with more updates containing little or only editorial changes.

We expect to get similar results in the future for other issuers with reasonably dense archiving history, such as Canada, Germany, Japan, and France, as for Australia and UK in terms of completeness, while issuers with even fewer snapshots or additional access barriers are likely to be less well covered. We can further deduce insights for our stated goal of building a multilingual global corpus with these results. We identify three main limitations for this goal that are not only technical properties of the Internet Archive but also reflect political economy, socio-technical, and international relations constraints:

1. Some states do not issue advisories at all or have started doing so much later. Others issue advisories with a far smaller scope. Although this is not an archive issue, it affects global analyses and makes some comparisons uneven by design.

2. Language and location visibility matter. English pages have a higher snapshot density on average. We observe this for Canada, where the English version of the index page is better represented than the French version. Similarly, locations such as Europe, the US and Japan seem to be better covered.

3. Access and capture can be shaped by blocking and geo-dependent hurdles. Archived replays return redirects or challenge pages that we cannot parse in some cases. This may occur for domains in specific countries and periods and is a known socio-technical constraint.



(a) UK vs. UK validation

(b) US vs. OSAC

(c) US vs. US index

(d) AU vs. AU index

Figure 4: Yearly coverage of Internet Archive-based advisories relative to (official) reference datasets. Each panel shows, for a given home country and year, the distribution of coverage scores across destination countries (violin plots with per-country points and yearly means).

Our findings support the use of the IAWM for large scale data collection. The IAWM is suitable for multilinguality and different issuer locations when it comes to European countries and Japan, but might lack data due to lower snapshot density for other countries either because of location or language (e.g., Hong Kong) together with often later start (e.g., Indonesia). It is still a good, reproducible, and transparent approach in the absence of viable alternatives.

6. Conclusion

We validated a transparent, reproducible, scalable, and language-independent pipeline to create time-series data for government-issued travel warnings using the Internet Archive's Wayback Machine as a data source. By validating our scraped corpus for the US, the UK, and Australia as major English-speaking countries against various reference sources, we showed that an Internet Archive-based reconstruction can achieve high coverage for some issuers and robust, though not exhaustive coverage for others. The main limitation resulting from our validation analysis for the three selected issuers arises for cases with very frequent updates, where the amount of archived snapshots does not always suffice to capture every version of every country page. At the same time, we did observe differences in terms of data availability of issuing governments of non-English speaking, non-Western-aligned countries for future extensions of the pipeline.

Despite these constraints, the resulting corpus already functions as a comparable resource across issuers and destinations, enabling longitudinal analyses of how states communicate risk and comparative studies of advisory levels and update dynamics. For many applications, it is more important to preserve relative differences and major changes over time than to observe every minor revision, and our corpus achieves this goal. In future work, we plan to extend the approach presented in this paper to include additional countries. Although we aim to create a geographically and linguistically balanced dataset, some of the limitations we outlined above may impede this goal to some extent. In addition, we plan to create additional variables that we want to extract from the advisory text using natural language processing methods, such as more granular

geographic locations and risk tags. We will make the data publicly available once we have a comprehensive and geographically and linguistically diverse corpus and dataset of government-issued travel advisories, which we envision to be valuable for qualitative and quantitative studies in various research areas.

7. Limitations

Our corpus construction relies entirely on the Internet Archive's Wayback Machine as a single archival source. Coverage in this archive is uneven across countries, time periods, and languages, and crawls can miss interim updates or complete pages because of robots.txt policies, technical outages, or crawl scheduling decisions. As a result, the data provide an approximation rather than an exhaustive record of all issued advisories. Our validation also depends on institutional reference data, such as the UK National Archives and the OSAC catalogue, which themselves have temporal gaps and are not guaranteed to offer perfect recall. Finally, we focus on English-language advisories from a small set of issuing countries with comparatively dense archival coverage in this paper, so the generality of our findings to other issuers, languages, and regions remains to be tested. However, we need to establish a baseline against which to validate less well covered issuing countries in the future, for which the countries chosen in this paper are ideal due to the geographic and linguistic reasons with regard to the IAWM outlined above.

8. Acknowledgements

We thank David Bencek, André Bluhm, Marius Hofmann and Daniel Racek, as well as the editors and two anonymous reviewers, for their helpful comments and feedback. We furthermore thank Jake Bickford and Natasha Kitcher from the UK National Archives for their support in obtaining the UK data for validation. The Center for Crisis Early Warning (Kompetenzzentrum Krisenfrüherkennung) is funded by the German Federal Ministry of Defense and the German Federal Foreign Office. The views and opinions expressed in this article are those of the author(s) and do not necessarily reflect the official policy or position of any agency of the German government.

9. References

Yasmin AlNoamany, Ahmed AlSum, Michele C. Weigle, and Michael L. Nelson. 2014. Who and what links to the Internet Archive. *International Journal on Digital Libraries*, 14(3-4):101–115.

Sanjay K. Arora, Yin Li, Jan Youtie, and Philip Shapira. 2016. Using the wayback machine to mine websites in the social sciences: A methodological resource. *Journal of the Association for Information Science and Technology*, 67(8):1904–1915.

Nicholas George Babey. 2019. The Politics of Travel Advisories: Foreign Policy and Error in Canada's Traveller Information Program. *The Journal of Intelligence, Conflict, and Warfare*, 2(1):15–33. Number: 1.

Anat Ben-David and Adam Amram. 2018. The Internet Archive and the socio-technical construction of historical facts. *Internet Histories*, 2(1-2):179–201.

Laura Braun and Christian Oswald. 2025. TRAVELWARN-Crawler: Constructing longitudinal datasets of government-issued travel warnings for political and social science research. In *Proceedings of the 21st Conference on Natural Language Processing (KONVENS 2025): Workshops*, pages 29–35, Hannover, Germany. HsH Applied Academics.

Yinxiao Chu, Xiaoyu Huang, and Tao Jin. 2021. Political relations and tourism: evidence from China. *Applied Economics*, 53(45):5281–5302. Publisher: Routledge _eprint: https://doi.org/10.1080/00036846.2021.1922591.

Oded Löwenheim. 2007. The Responsibility to Responsibilize: Foreign Offices and the Issuing of Travel Warnings. *International Political Sociology*, 1(3):203–221.

Jamie Murphy, Noor Hazarina Hashim, and Peter O'Connor. 2007. Take Me Back: Validating the Wayback Machine. *Journal of Computer-Mediated Communication*, 13(1):60–75.

Richard Sharpley, Julia Sharpley, and John Adams. 1996. Travel advice or trade embargo? The impacts and implications of official travel advice. *Tourism Management*, 17(1):1–7.

Mike Thelwall and Liwen Vaughan. 2004. A fair history of the Web? Examining country balance in the Internet Archive. *Library & Information Science Research*, 26(2):162–176.

Gerhard Weikum, Nikos Ntarmos, Marc Spaniol, Peter Triantafillou, András Benczúr, Scott Kirkpatrick, Philippe Rigaux, and Mark Williamson. 2011. Longitudinal Analytics on Web Archive Data: It's About Time! pages 109–202.

A. Supplementary Material

Table 3: Country–issuer URLs

Country	Time-start	URL	# all	# 200
Australia	1997	http://www.dfat.gov.au/consular/advice/advices_mnu.html	153	47
	1998	http://www.dfat.gov.au/consular/advice/consadvice_main.html	31	7
	1998	http://www.dfat.gov.au/consular/advice/index.html	224	38
	2004	http://www.smartraveller.gov.au/zw-cgi/view/Advice/	1463	991
	2015	http://smartraveller.gov.au/countries/list.html	47	14
	2016	http://smartraveller.gov.au/Countries/Pages/list.aspx	172	65
	2019	https://www.smartraveller.gov.au/destinations	545	513
		Total	**2635**	**1675**
Canada (en)	1997	http://www.dfait-maeci.gc.ca/graphics/cosmos/CNTRY_E.htm	139	59
	2000	http://voyage.dfait-maeci.gc.ca/destinations/menu_e.htm	145	62
	2003	http://www.voyage.gc.ca/dest/ctry/reportpage-en.asp	1057	772
	2008	http://www.voyage.gc.ca/countries_pays/menu-eng.asp	581	159
	2012	http://travel.gc.ca/travelling/advisories	4699	3850
		Total	**6621**	**4902**
Canada (fr)	1997	http://www.dfait-maeci.gc.ca/graphics/cosmos/cntry_f.htm	42	22
	1999	http://www.dfait-maeci.gc.ca/travelreport/menu_f.htm	93	14
	2003	http://www.voyage.gc.ca/dest/ctry/reportpage-fr.asp	398	267
	2009	http://www.voyage.gc.ca/countries_pays/menu-fra.asp	444	109
	2012	http://voyage.gc.ca/voyager/avertissements	1908	1357
		Total	**2885**	**1769**
China	2005	http://www.fmprc.gov.cn/chn/lsfw/lsxw/fbfgjhcszysx/default.htm	42	36
	2012	http://cs.mfa.gov.cn/lsyj/gbtx/	5	5
	2013	https://cs.mfa.gov.cn/gyls/lsgz/lsyj/	600	466
		Total	**647**	**507**
France	2000	http://www.dfae.diplomatie.fr/voyageurs/etrangers/avis/conseils/minute.asp	38	38
	2005	http://www.diplomatie.gouv.fr/fr/conseils-aux-voyageurs_909/index.html	850	684

Continued on next page

Country	Time-start	URL	# all	# 200
	2006	http://www.diplomatie.gouv.fr/fr/conseils-aux-voyageurs	2167	1906
		Total	**3055**	**2628**
Germany	1998	http://www.auswaertiges-amt.de/5_laende/index.htm	160	65
	2000	http://www.auswaertiges-amt.de/www/de/laenderinfos/reise_warnung_html	281	161
	2006	http://www.auswaertiges-amt.de/diplo/de/Laenderinformationen/01-Reisewarnungen-Liste.html	187	162
	2006	http://www.auswaertiges-amt.de/diplo/de/Laenderinformationen/01-Reisewarnungen.html	63	32
	2010	http://www.auswaertiges-amt.de/DE/Laenderinformationen/01-Reisewarnungen-Liste_node.html	250	136
	2017	https://www.auswaertiges-amt.de/de/ReiseUndSicherheit/10.2.8Reisewarnungen	1992	1296
	2024	https://www.auswaertiges-amt.de/de/reiseundsicherheit/10-2-8reisewarnungen	63	53
		Total	**2996**	**1905**
Hong Kong (ch)	2021	https://www.sb.gov.hk/chi/ota/index.html	37	33
		Total	**37**	**33**
Hong Kong (en)	2009	https://www.sb.gov.hk/eng/ota/	486	256
		Total	**486**	**256**
India	2016	https://www.mea.gov.in/travel-advisories.htm	254	241
		Total	**254**	**241**
Indonesia	2017	https://safetravel.kemlu.go.id/	49	46
		Total	**49**	**46**
Japan	2003	http://www.anzen.mofa.go.jp/	6899	2992
		Total	**6899**	**2992**
Mexico	2003	http://www.sre.gob.mx/delviajero/	255	180
	2024	https://portales.sre.gob.mx/guiadeviaje/	13	13
		Total	**268**	**193**
Russia	2016	http://www.mid.ru/ru/preduprezdenie-dla-rossijskih-grazdan1	125	107
	2023	https://www.mid.ru/ru/useful_information/information/preduprezhdeniya_dlya_rossiyskikh_grazhdan/	2	2
		Total	**127**	**109**
SAR Macao	2023	https://www.dst.gov.mo/zh-hant/tourism-crisis-management/tourism-crisis-management-travel-alert.html	13	13
		Total	**13**	**13**
United Kingdom	1997	http://www.fco.gov.uk:80/reference/travel_advice/countries.html	56	26

Continued on next page

Country	Time-start	URL	# all	# 200
	1998	http://193.114.50.10/travel/	44	15
	2002	http://www.fco.gov.uk/servlet/Front?pagename=OpenMarket/Xcelerate/ShowPage&c=Page&cid=1007029390590	764	574
	2008	http://www.fco.gov.uk/en/travelling-and-living-overseas/travel-advice-by-country/	397	179
	2009	http://www.fco.gov.uk/en/travel-and-living-abroad/travel-advice-by-country/	588	345
	2013	https://www.gov.uk/foreign-travel-advice	11382	9699
		Total	**13231**	**10838**
United States	1996	http://travel.state.gov/travel_warnings.html	4970	1766
	2004	http://travel.state.gov/travel/cis_pa_tw/tw/tw_1764.html	6874	1672
	2004	http://travel.state.gov/travel/warnings_current.html	534	112
	2014	http://travel.state.gov/content/passports/english/alertswarnings.html	4110	463
	2015	https://travel.state.gov/content/passports/en/alertswarnings.html	5129	527
	2017	https://travel.state.gov/content/travel/en/traveladvisories/traveladvisories.html	19198	8647
	2025	https://travel.state.gov/en/international-travel/travel-advisories.html	156	155
		Total	**40971**	**13342**

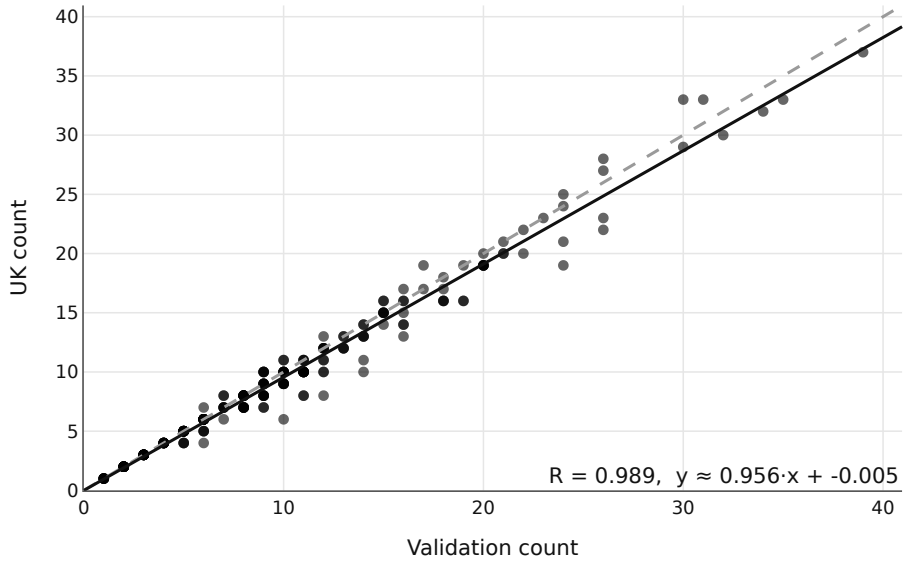


Figure 5: Agreement Internet Archive UK vs. The National Archives UK for the year 2022

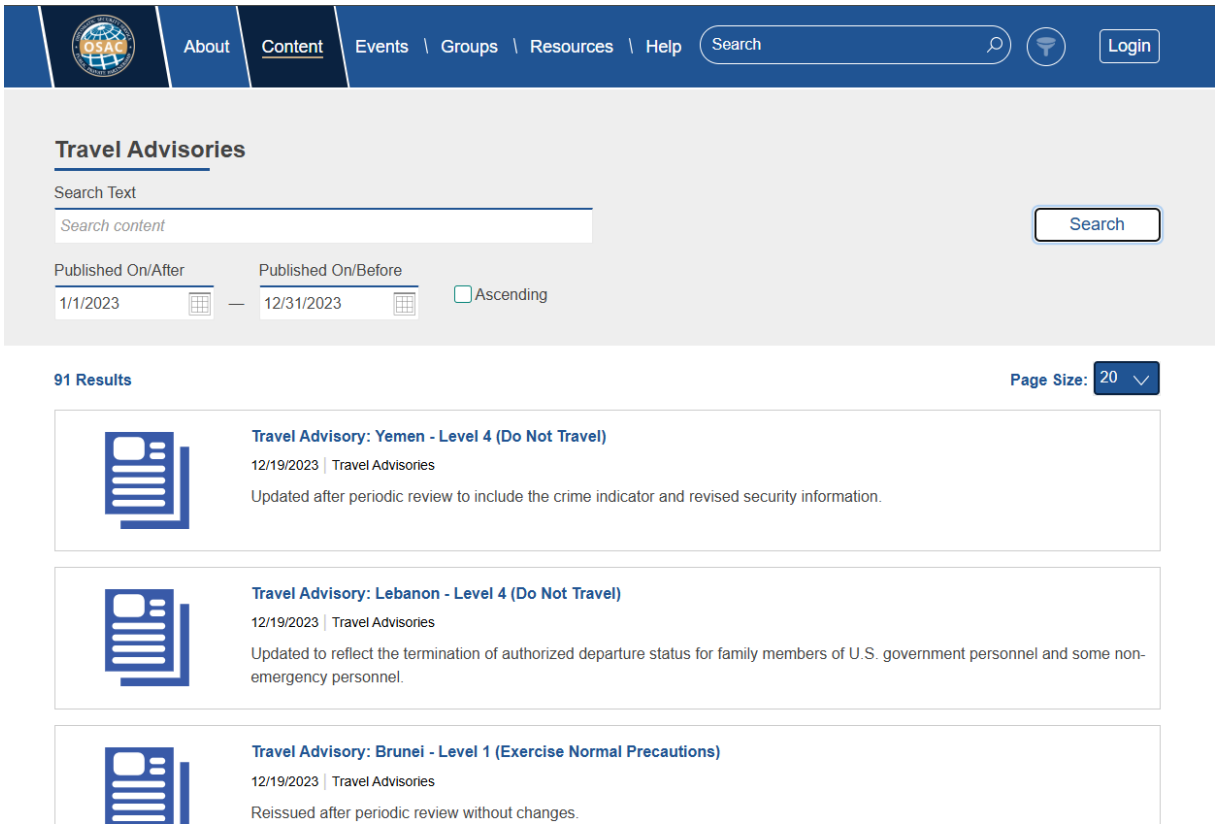


Figure 6: OSAC catalogue with filter on year 2023

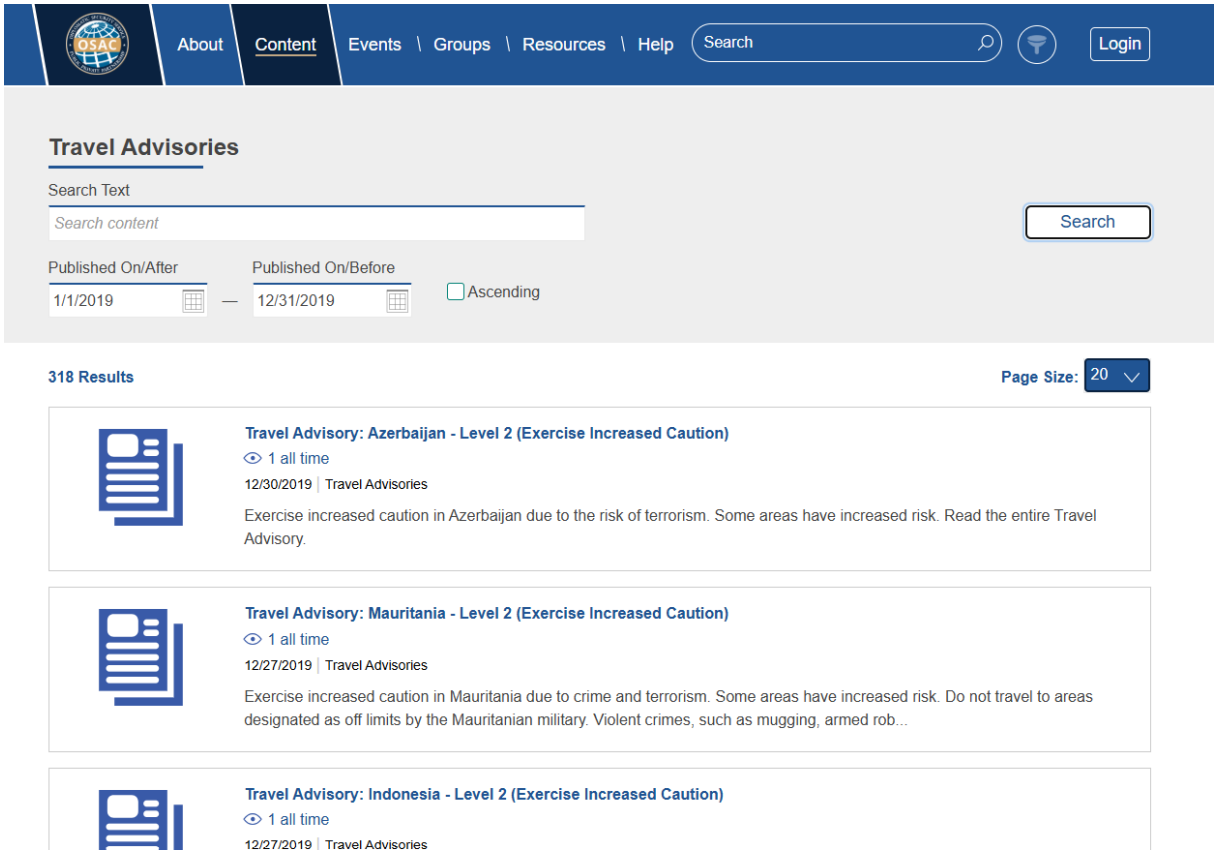
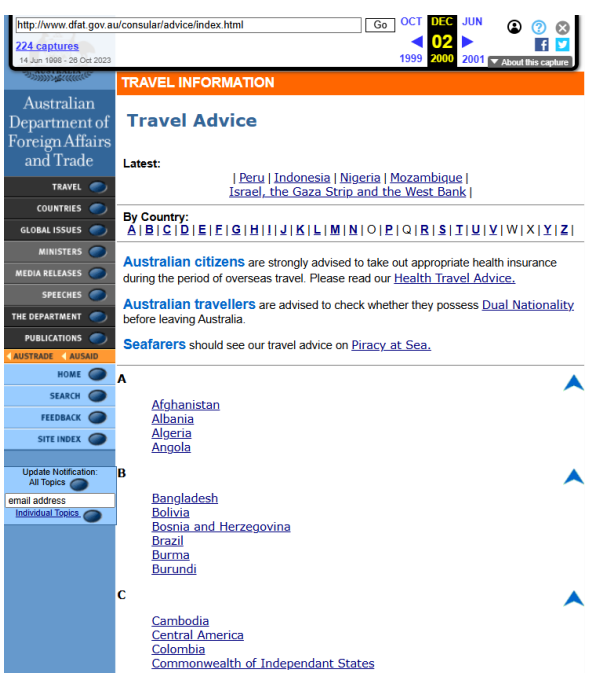
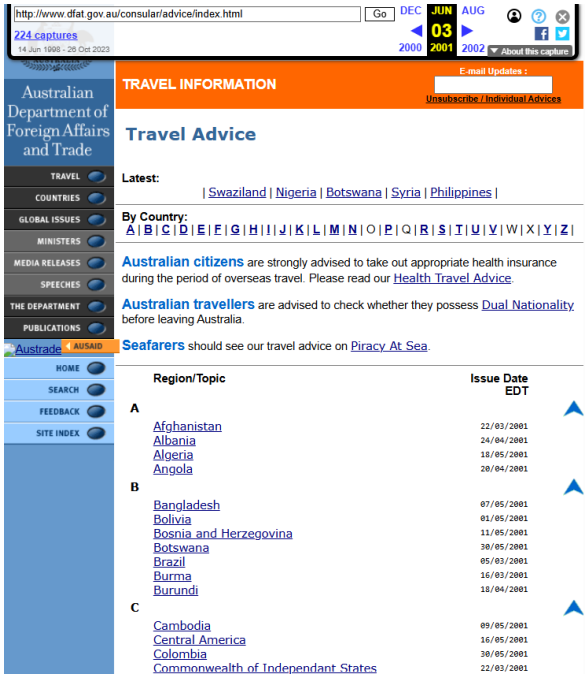


Figure 7: OSAC catalogue with filter on year 2019



(a) Australia advisory index page (without date)



(b) Australia advisory index page (with date)

Figure 8: Australia advisory index pages before and after the design change