

Computer-assisted research and the construction of a "dried fish literature"

Eric Thrift with Derek Johnson, Ben Belton, and Jonah Olsen

Three years ago, we set out on what appeared, at the time, to be a fairly straightforward task: writing a survey of the published literature on dried fish in South and Southeast Asia. As we struggled to make sense of hundreds of references to dried, fermented, salted, and smoked fish returned by Google Scholar's opaque algorithms, we found that technology had come to play a governing role in our unexpectedly formidable project. Instead of simply reading and summarizing an established body of scholarship, we were actively working to define a "dried fish literature" through the intermediary of Google Scholar, Zotero, and additional tools of our own design. In this essay, I reflect on how communicative tools and technologies shaped our reading of this literature, structuring our ways of seeing and interacting with "dried fish" as a category of knowledge.

THE PROJECT

Our project began in 2018 as a relatively modest literature review concerning dried fish in South and Southeast Asia. Over the course of the next three years, this work expanded into the much larger project of building and analyzing a comprehensive, global bibliographic database of publications on dried fish. By the end of this activity we had collected and catalogued over 3000 tagged items pertaining to dried fish, including a core set of 1130 publications we identified as directly relevant to dried fish and for which we had collected full bibliographic details (Dried Fish Matters 2021). In the process of this work, we also created an online tool for generating graph visualizations from our dataset, and an associated command-line tool developed for querying and updating the bibliographic database (Thrift 2021b; 2022). We presented our methods and analysis of this database in a paper at the MARE Conference in 2019 (Johnson et al. 2019), a training module on literature reviews for an online course on transdisciplinarity in fisheries research (Johnson and Thrift 2020), and a 135-minute webinar (Thrift, Belton, and Johnson 2020). Our overall analysis of the literature database was written up as a 15,000-word manuscript submitted to *Fish and Fisheries*, accompanied by five appendix documents containing supporting information about the dataset and our analytic methods (Belton et al. Forthcoming). The database has subsequently been used in the preparation of a chapter in an FAO Technical Paper on small fish (Johnson et al. Forthcoming), and continues to support and integrate the work of our students and research partners.

Our literature review was originally conceived as the first major research output of the SSHRC-funded partnership grant "Dried Fish Matters: Mapping the social economy of dried fish in South and Southeast Asia for enhanced wellbeing and nutrition". Initially, we thought the literature review would be a fairly straightforward endeavour. We had already argued, in our project proposal, that research concerning dried fish was sparse, but claimed that the Project Director and other members of the research teams had been "major contributors" to the "small body of research in this area". We envisaged that a student research assistant would be able to

assemble a fairly comprehensive set of publications on dried fish through a scholarly database search, which we could then classify and summarize according to the themes of our own project. This literature review would then guide our own research during our seven-year partnership grant, by helping us to identify knowledge gaps worthy of being filled by the various research partners and teams. Summarizing our intentions in a planning document at the inception of this project, we wrote:

The intent of the paper is to establish a baseline review of knowledge on dried fish globally, but with focus particularly on Asia and Africa, as those are the areas with the largest dependent populations and with the most pressing nutritional needs. The paper will also provide a valuable framework, identifying major gaps in knowledge, for DFM. It will reinforce DFM's *raison d'être*, and provide an empirical basis to position our efforts.¹

Our initial plan was to use a Dropbox folder to share resources previously collected by Ben Belton, Derek Johnson, and if possible other colleagues, to be extended through an annotated bibliography generated by a research assistant through a systematic literature search. We would then spend about two months analyzing the literature for themes and gaps, followed by another three months writing. We estimated that the entire project, start to finish, would take six months.

At the outset, we noted that our intention was to combine a search for literature explicitly concerning dried fish with a kind of “forensic analysis” of the broader fisheries literature, looking for incidental mentions of dried fish. (At this point we still considered the “dried fish literature” to be a subset of the “general fisheries literature”, though we later came to make the claim that they are substantially distinct.) We acknowledged that our review would be guided by a set of pre-established themes derived from our “existing knowledge of the dried fish literature”, to be adjusted based on our actual findings. This prior knowledge drove our understanding of dried fish as having historically been central to coastal economies and to food systems globally, coupled with the view that the published literature on dried fish was fragmented and partial. We estimated that the literature on dried fish was dominated by reports on the results of technical analyses of products or processing practice; confirming this hypothesis should, we thought, bring justification to our broader research project.

We began the literature review with an established set of inclusion criteria for processed fish products, corresponding roughly to *dried or fermented but not canned fish*. We came to acknowledge a “distinctive definition of dried fish” that was not entirely “dry”:

During a series of project inception meetings in June 2019, project participants agreed to an expansive definition of dried fish: any processed fish that can be stored without artificial cooling. As such, dried fish includes sun-dried, smoked, salted, and fermented products. We include fish sauce in our definition, but not canned fish. Dried fish can be eaten whole or, more commonly, is integrated

¹ “Summary notes of Dried Fish Review paper brainstorming”, 2018-03-21, internal document DFM_MEM_Dried-fish-lit-review-notes_2018-03.

into food preparations as a component or condiment to enhance flavor and nutrition. (Belton et al. Forthcoming)

As suggested by the statement above, recognizing that fermented fish actually had much of the same role as dried fish in the Southeast Asian sites included in our research project, our broader intention was to describe *dried fish* in social-economic terms rather than with reference to its physical characteristics. In this sense, our study addressed several claims about dried fish that were fundamentally definitional insofar as they constructed *dried fish* as a conceptual category, rather than as a material product:

- Dried fish is produced in value chains with four levels of actors: producers, processors, wholesalers, and retailers.
- Dried fish is generally processed by women.
- Dried fish meets the food security needs of vulnerable people.
- Dried fish is highly nutritious.
- Dried fish has important cultural value.
- Dried fish has low food quality and safety characteristics, from the perspective of a large proportion of current “technical” research.

We expected that our literature review would refine these claims, for example by identifying specific evidence of nutritional or cultural value, or by providing quantitative support to claims about the role of women or the dominant trends in technical research. While our study unsurprisingly achieved this broad objective, it is noteworthy that our methods and tools influenced the specific findings we were able to advance – including the selection of evidence for each of these claims, and the ways we were able to present that evidence.

GOOGLE SCHOLAR: QUERIES AND KEYWORDS

We opted to use Google Scholar to locate publications for the literature review sample, due to the search engine’s strong coverage of grey literature – including technical reports from the FAO and other intergovernmental organizations – which we knew to be the only sources of published information on dried fish for some of the locations included in our project. Studies comparing Google Scholar to “controlled” academic publication indexes, such as Web of Science and Scopus, have confirmed that Google Scholar returns a significantly larger number of citations, since approximately half its sources are from non-journal sources (including conference proceedings) and, for some searches, more than one in three are from non-English publications (Martín-Martín et al. 2018). Google Scholar’s coverage appears to be stronger for some disciplines than others, providing lower-quality results in some fields but higher citation coverage of the social sciences and humanities than controlled indexes (Halevi, Moed, and Bar-Ilan 2017). The most current comparisons suggest that Google Scholar locates 90% of citations found in Scopus and 94% of those found in Web of Science; conversely, Scopus and Web of Science respectively only locate 58% and 55% of results returned by Google Scholar (Martín-Martín et al. 2021). These characteristics made Google Scholar an attractive choice for a literature review intended to focus on fragmented research often located in “low-ranking” publications.

Jonah Olsen, then an anthropology student at the University of Manitoba, was hired to take on a literature search in May and June of 2018. We established that Jonah would conduct searches combining terms for dried fish either with names of areas of geographic focus within our project, or with keywords related to the themes of our research. For the geographic regions, we settled on the six countries of our project (keywords “India”, “Sri Lanka”, “Bangladesh”, “Cambodia”, “Myanmar” / “Burma”, and “Thailand”), sub-national administrative units in which our research terms were located (“Andhra Pradesh”, “Gujarat”, etc.), and regional units (“Asia”, “South Asia”, “Southeast Asia”). For the thematic keywords, we chose eleven terms that reflected the major focus areas of our transdisciplinary project: economics (“demand”, “value chains”, “economics”), health (“health”, “food security”, “nutrition”), social science (“gender”, “women”, “wellbeing”, “ethnography”), and technical interventions (“technology”). For dried fish, we opted to include the keywords “dried fish” and its variant “dry fish”, “cured fish”, “smoked fish”, and “fermented fish”, reflecting the working definition advanced in our project proposal of *dried fish* as any product “preserved using simple techniques including sun drying, salting, fermentation, and smoking”.

These search terms produced a total of 110 combinations, each of which was entered into Google Scholar as an individual search query. Using the browser plugin for the Zotero reference manager, we retrieved the metadata and PDF for publications from each consecutive search until we judged that the items being returned were no longer relevant, then reviewed all results to remove false matches for the query. Each search resulted in a median of seven valid results, with a low of one and a high of 43 (fig. 1). In general, we obtained the fewest valid results from thematic keyword searches, and the most valid results for searches with geographic keywords identifying sites of major production (“India”, “Sri Lanka”, “Kerala”, “Cambodia”, “Bangladesh”, “Indonesia”, and “Southeast Asia”). Not all results deemed “valid” necessarily had a direct correlation to the specific search term as used, but were sometimes retained due to their general or incidental importance to dried fish. In some cases, a result item incidentally referenced a country identified in the search term but, in relation to dried fish, directly concerned a different country altogether. Several result items surfaced through more than one search; these items were imported repeatedly for each search then merged within Zotero. At the end of the first search round (May-June 2018), we were left with 811 items. 443 of these items were only incidentally relevant to dried fish, and 367 were written from a “technical” or food science perspective (104 items overlapped these two categories). This left only 105 items that addressed dried fish from the major themes of interest that we had identified.

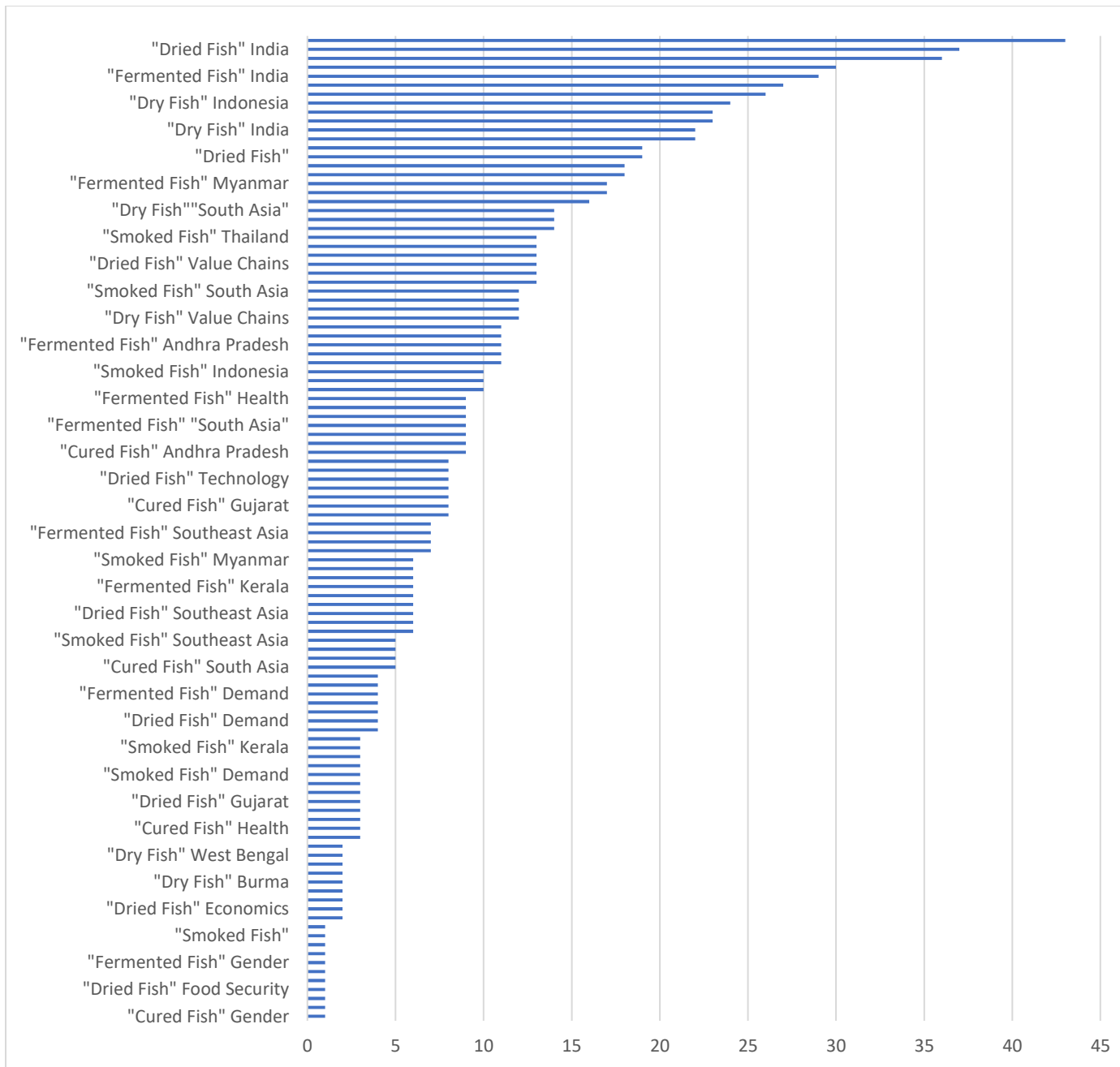


FIGURE 1. Number of valid results returned by each search query in the first round of Google Scholar searches. Each bar represents a single search query. Labels represent the text of search queries used in Google Scholar; in the interest of legibility, not all search terms are shown.

ZOTERO AS QDA PLATFORM

We chose Zotero to manage our search results, as an open source reference management platform that provides a web browser plugin to facilitate downloading publications and their metadata from scholarly search engines, libraries, and publication websites. Billed as “a free,

easy-to-use tool to help you collect, organize, cite, and share research” (Corporation for Digital Scholarship 2022), Zotero was originally developed at the Roy Rosenzweig Center for History and New Media at George Mellon University and is now managed by the non-profit Corporation for Digital Scholarship, as part of its goal to provide sustainably managed, open-source infrastructure for digital humanities scholarship. Zotero was initially built as a plugin for the Firefox web browser, but now exists as a standalone desktop application and a web-based service, with an optional subscription model for online data storage. As it was first developed as a citation management tool, Zotero has many features in common with tools such as EndNote, Mendeley, and Refworks – including batch import of references from online databases, integration with word processors, and creation of bibliographies – but distinguishes itself as the only major open-source reference management platform (Ivey and Crum 2018). Although we have seen some publications regarding creative uses of Zotero, for example as an information literacy teaching tool (Winslow, Skripsky, and Kelly 2016), we have not encountered recent discussion in academic discourse of how Zotero itself shapes the construction of knowledge.

We were drawn to Zotero’s collaborative functions, which allow teams of researchers to synchronize sets of publications, metadata, notes, and tags using a cross-platform desktop application. Zotero provides users the option to organize items using arbitrary tags, which persist as items are moved or copied to different places in the library, and collections (working folders), which gather sets of items but are not permanently associated with the items they contain. We were particularly interested in possibility of using Zotero tags for systematic analysis, by filtering publications according to individual tags or tag combinations. We further anticipated that, beyond supporting our literature review as an immediate goal, the Zotero library itself would serve as a valuable research output – offering a catalogue of references that could be queried, analyzed, and cited by members of the Dried Fish Matters partnership and beyond. Anyone with access to the library and a word processor plugin is able to insert citations to publications concerning dried fish in a document written in Word, Open Document (LibreOffice), or Google Docs formats, and to generate a correctly formatted bibliography, thanks to the word processor plugins that are packaged with Zotero.

Once the search results had been collected in our database, we divided up the library alphabetically and coded each item using a predetermined list of tags. As many of the items were automatically assigned subject tags by the software on import, drawing on the publisher-supplied keywords and subject headings, we used the hash symbol (#) as a prefix to distinguish our analytic codes from any other tags that might be present in the database. In our first round of analysis we first went through the library distinguishing between items that directly and incidentally concerned dried fish, using the tags [#RELEVANCE: Direct] and [#RELEVANCE: Indirect], and assigning one of twelve thematic tags derived from the project’s core topics. Items that related to value chains or microeconomics were also assigned a segment tag, and those that concerned the primary production segment (i.e., fishing) were assigned the tag [#FISHERY: Inland] or [#FISHERY: Coastal] (fig.2). Following this process, we read through the library a second time to review and validate the tagging, wherever possible working with a set of items that another member of the team had tagged in the previous round.



FIGURE 2. Tagging tree used in our quantitative analysis of the dried fish literature (Johnson et al. 2019).

Our thematic categories were refined as we became more familiar with the literature and saw the need to add, merge, or split tags. This process is consistent with the practices of qualitative data analysis (QDA), as employed using tools such as Atlas.ti or NVivo, with the difference that Zotero tags (codes) are necessarily associated with documents rather than individual segments within a document – such as passages within an interview transcript – as would typically be done in QDA.² In terms of analytic functions, Zotero provides similar coding options to typical QDA software, including the ability to combine and rename tags, filter by tag or tag combination, and create reports of items that match specific tags or other criteria. Zotero “notes”, which can be associated with items or created as standalone documents, mirror the “memos” of QDA software. Although there is no function to assign definitions directly to tags in Zotero, standalone notes can be used for that purpose, as they may be associated with the tags they describe.³ Similar to the reporting features of QDA software, Zotero’s “saved search” function allowed us to create dynamically-updated folders listing specific combinations of tags, using Boolean expressions: for example, the “Dried Fish Sample” was constructed using a tag query

² The current beta version of the Zotero desktop client includes a built-in PDF reader, which allows users to associate annotations and tags with individual passages in a stored PDF document stored within Zotero. As of this writing, this feature is incomplete: the Zotero client does not provide a mechanism for searching or filtering tags associated with document passages, or for searching annotations across multiple documents; nor does it accommodate the annotation of items within shared libraries.

³ QDA tools such as Atlas.ti, which are designed to support grounded research, help researchers write texts within the software using memos that are linked to quoted passages. Zotero accommodates a comparable workflow: notes can be both tagged and linked to other related items, including other notes; the development version of Zotero also allows notes to embed citations to items in the bibliography, and for notes to be imported directly into a word processor along with their linked citations.

corresponding to [#RELEVANCE: Direct] AND NOT [#exclude], while the aggregate tag [##THEME: food science] was created through a query corresponding to the expression [#THEME: food engineering] OR [#THEME: food chemistry and microbiology] OR [#THEME: food safety]. Beyond these similarities, Zotero offers more robust methods for documenting and analyzing structured metadata than QDA software. While the most significant metadata are captured in basic bibliographic fields – author, title, year of publication, and publication type – Zotero provides a total of 125 distinct metadata fields, allowing documents to be catalogued, sorted, and searched by creator type, conference title, and the like (Corporation for Digital Scholarship 2020).

An area in which Zotero proved unexpectedly useful was in facilitating a collaborative workflow. During the course of our thematic tagging, we adopted a number of procedural tags, intended to communicate messages to other members of the teams during their subsequent review of the library items. For example, we used an exclamation mark [!] to indicate a reference of particular interest to the team; tags prefixed with the commercial at symbol [@] to indicate that a reference had been manually reviewed, cited in another context [@cited-WG2], or assigned to a particular reviewer's queue [@EDT]; and question mark-prefixed tags such as [?geo], [?theme], or [?noPDF] as temporary codes to indicate when the person tagging was uncertain about a particular tag, or to flag items to discuss with the author group for a collective decision.

CHANGING SCOPE: GOOGLE SCHOLAR AND SAMPLING

Our decision to assemble and describe a *dried fish literature* involved a change in scope, from our original goal of analyzing contemporary literature by country in South and Southeast Asia, in favour of a global thematic study with attention to geographic coverage. This shift was partly a response to feedback from the SSHRC Expert Panel that had reviewed our project application, and that had recommended a more global and historically informed study:

The panel noted that, with rare exception, the literature review is limited to sources in the 21st century. This is understandable given the relative ease of accessing current literature electronically. As the project unfolds, literature from the dark ages (i.e., before the internet) might well reveal additional material of historical importance.

The panel also suggests the literature review be expanded beyond the regions in question. For example, the cod fisheries of Newfoundland were driven originally by production and trade of dried fish, something that continued over well into the 20th century. Dried fish is an important product in other parts of the world as well. Without diluting the focus on South and Southeast Asia, the researchers might benefit from situating these regions in a larger global history of dried fish.⁴

The recommendation to consider an expanded scope for this project did not directly inform our design of a reference management workflow using Google Scholar and Zotero. We quickly discovered, however, that these tools would facilitate – and indeed become indispensable to – a comprehensive review. Due to the facility with which they allowed us to collect and analyze a large volume of publications, these tools encouraged us to set a goal of creating a census of the

⁴ SSHRC Expert Panel Report, Partnership Grant Application number 895-2018-1017, 2017-12-12.

global dried fish literature. Had we limited our review, as we might have done fifteen or twenty years ago, to the publications we could retrieve through our local library, the shape of this study would have been entirely different. In the absence of a full-text search database (Google Scholar), we would have been forced to rely on the subject listings available in our library catalogue, which would have produced only a handful of references: recreating a sample based on the publisher-supplied keywords extracted by Zotero for each item in our library, we would have obtained a sample of only 235 publications in a search for keywords that partially match the terms "dry" OR "dried" AND "fish" (Table 1). Inclusion of grey-literature sources would have been far more tedious, as it would have required consulting multiple indexes for periodicals and government or intergovernmental agency reports. Limiting ourselves to academic articles and books, we would have ignored one in ten of the publications from our final collection – which includes 53 conference papers and presentations, 12 patents, 7 manuscripts and unclassified documents, 2 newspaper articles, 4 magazine articles, 53 technical reports, and 28 theses. In the end, we might have been satisfied to combine a “proximity sample” of publications from our local library with a “convenience sample”, such as the Dropbox folder of 93 items assembled at the beginning of this project – mainly collected during the course of Ben Belton’s own research in Bangladesh and Southeast Asia, and reflecting the thematic and methodological biases of that origin.

keyword	items
dried cod	1
dried fish	180
dried fish products	2
dried fish tourism	2
dried fish training abroad	1
dried flying fish	1
dried salt-cured cod	1
dried salted coalfish	1
dried salted fish	1
dried-fish	1
dried-salted fish	1
dry fish	31
dry fish consumption	2
dry fish export	1
dry fish industry	2
dry fish market	1
dry-curing fish	2
drying of fish	1
fish drying	3
TOTAL	235

TABLE 1. Number of items matching publisher-supplied subject keywords containing the terms “dried fish” or “dry fish”, within all sources returned through Google Scholar searches. Keywords are mutually exclusive.

To accommodate the broadened scope of our literature review, we conducted a second round of search queries in October 2019, adding a further 1016 items to the Zotero library. Our primary search queries employed the phrases “cured fish”, “dried fish”, “dry fish”, “fermented fish”, “smoked fish”, “salt fish”, and “salted fish” in combination with country-level geographic

descriptors taken from the list of countries and territories recognized by the United Nations, extended by several regional sub-country descriptors identified by the researchers as of likely importance (e.g., “Indian Ocean”, “Newfoundland”). Supplementary searches were then conducted using product terms that emerged during the initial review, in an effort to capture publications that used local instead of generic terms for dried and fermented fish: “Bacalao” / “Bacalhau”, “Klipfisk” / “Klippfisk”, “Lutefisk”, “Stockfish”, and “Surstromming”. Unique results generated by variant terms (e.g., “UK” for “United Kingdom”) were consolidated into a normalized query tag referencing the primary term. Where a publication was retrieved by more than one search, all applicable search queries were applied to that item.

Across the two rounds of searches, 553 queries yielded valid results, with a median of 5 valid results per query (fig.3). Of the searches that produced the greatest number of results, the top five reproduced queries that had been present in the first search round (four of the five related to dried fish in India and Bangladesh); queries that yielded 20 or more items included the search terms including geographic keywords for Korea, the Caribbean, Brazil, Africa, the Philippines, and Zambia (Table 2).

<i>"Salted Fish"</i>	79
<i>"Cured Fish" India</i>	46
<i>"Dried Fish" India</i>	41
<i>"Dried Fish" Bangladesh</i>	33
<i>"Fermented Fish" India</i>	32
"Fermented Fish" Korea	32
"Smoked Fish" Caribbean	32
<i>"Dried Fish" Andhra Pradesh</i>	31
"Dry Fish" Caribbean	31
"Salted Fish" Brazil	26

TABLE 2. Search queries and number of valid results, rounds 1 and 2 combined. Round 1 searches are italicized.

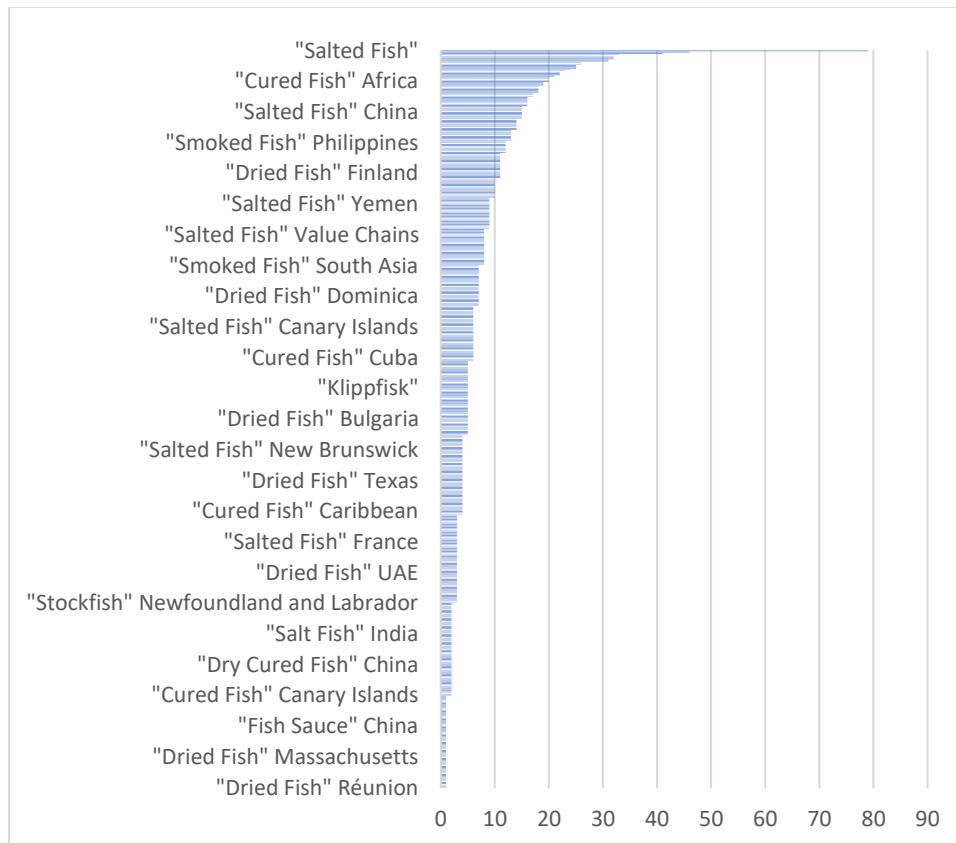


FIGURE 3. Number of valid search results per query, round 2 (553 queries total).

While we are confident that, after several rounds of searches, our Zotero library constituted a reasonably complete representation of the dried fish literature, our reliance on the proprietary crawling and search ranking algorithms of Google Scholar contributed to a result set that was not entirely within our control. The lack of transparency in Google Scholar's indexing and result ranking methods has been flagged by some researchers as a serious limitation, and an obstacle to the use of the search engine for rigorous purposes of systematic reviews or bibliometric analysis (Halevi, Moed, and Bar-Ilan 2017). We know from Google's own instructions for publishers that the indexing of sources in Google Scholar is intended to be comprehensive, by including any publication that has a distinct metadata (description) page that is publicly accessible on the web containing an abstract. Google Scholar recognizes documents as scholarly publications either due to the presence of publisher-supplied metadata tags identifying the publication as a journal article, or through a combination of heuristic measures, including the layout of the document (i.e., a title at the top of the document, followed by authors and affiliations, and inclusion of a references or bibliography section), and hosting on a University web server or on another site using repository management software. The inclusion and ranking of documents in search results is influenced by the availability of a full-text version of the document accessible to Google, if the search keywords are not found within the document abstract; and by other factors including the number of inbound citations (Google n.d.).

By following the search strategy described above, we became dependent on a census/sampling methodology that was linked to Google Scholar, and whose methodological rigour was effectively defined in relation to the opaque operations – and limitations – of that search engine. We were able to include a reasonably wide range of sources, but not all pre-digital (“dark ages”) or print-only resources from beyond the digital divide, let alone non-English-language publications. In our final sample, the median date of publication was 2008, or 10 years prior to the date of our first web searches; three quarters of our items were published in the year 2000 or later, and three-quarters of the remaining items were published in 1980 or later.⁵

Many relevant non-digital publications were indexed by Google Scholar, and returned in search results, via AGRIS, a bibliographic catalogue managed by the FAO. AGRIS aggregates data on publications supplied by over 15,000 organizations, including academic libraries, research institutes, government agencies, publishing houses, private foundations, and international organizations. The index itself has operated since 1974 and includes “books, journal articles, monographs, book chapters, datasets and grey literature – including unpublished scientific and technical reports, theses, dissertations and conference papers in the area of food and agriculture” (Food and Agriculture Organization of the United Nations 2022). Most of the items in this database do not provide full-text links; many refer to non-English sources; and quite a few entries point to older, print-only sources that are available in the catalogues of data provider institutions. Our Zotero library includes 166 items retrieved from the AGRIS database, 118 of which we excluded from the dried fish sample due to the unavailability of the source material (including instances in which the listed item was an unpublished conference paper), lack of bibliographic detail other than the author and title, or publication in a non-English-language source. 80% of the items returned from AGRIS belonged to the “technical” category.

TAGGING

Google Scholar, like any other contemporary catalogue of scholarly publications, allowed us to retrieve a much greater number of results based on matching terms from their full-text content. But it was the Zotero reference manager that allowed us to combine the results of multiple searches to support large-scale analysis, collecting the abstract and metadata for each publication and a copy of the text itself.

The original plan for the literature review was to limit our attention to the countries included within the Dried Fish Matters project, focusing on two major axes: country addressed (India, Bangladesh, Indonesia, Myanmar, Cambodia, Sri Lanka, and Thailand), and thematic coverage, making a broad distinction between “technical” literature – primarily associated with the disciplines of biology, nutrition, and food science – and a set of smaller themes associated with history and anthropology, economics, ecology, and policy and development studies. As the geographic scope of our literature review expanded, we were forced to explore alternatives for

⁵ To counteract the regional bias, we invited our partners to incorporate scans of local sources into dedicated subfolders in the Zotero database, particularly theses, policy documents and legal texts, technical reports, and other primary or grey literature sources. Although we received some submissions, the process proved labour-intensive and technically challenging for some of our collaborators; in several cases, our partners were unable to locate any local publications that addressed dried fish.

simplifying our approach. In one of the earliest iterations of our tagging strategy, we developed an analytic frame that would combine four axes of interest – theme, region, date, and fishery type – which would have involved 72 tag combination permutations (Table 3). In later iterations we chose to remove the date and fishery variables, making room for product type and value chain segment variables, and increased the number of top-level regions to 11 while reducing the top-level themes to 5.

Theme	Region	Date	Fishery
#THEME: ecology #THEME: economy #THEME: gender #THEME: nutrition and food security #THEME: policy and governance #THEME: value chain #THEME: well-being	#REGION: South Asia #REGION: SE Asia	#DATE: Contemporary #DATE: 20th Century #DATE: pre-20th Century	#FISHERY: coastal #FISHERY: inland

TABLE 3. Original analytic frame for the literature review, showing prefixed Zotero tags.

Most of the thematic tags used in our final analysis relate to areas of particular interest for our own work. [Cooking and food] relates closely to the broad [Culture, social relations and well-being] tag, and to our original search term “consumption”, but highlights the importance of food as a symbolic marker. [Gender] is also a sub-tag of the [culture, social relations, and well-being] tag, and reflects our assessment that gendered is profoundly important in the analysis of social economies of dried fish and often an important axis of distributional inequalities. [Gender] and [labour] cut across the value chain, and are important points of reference in normatively evaluating its functioning. The [economy] tag refers to broad, macro-economic analyses of the place of dried fish in economic systems; it marks references that speak to the broad economic context within which specific dried fish value chains operate. The [history and change] tag includes two analytical dimensions: references that are overtly historical in their orientation, in the sense looking at patterns of dried fish use in the past, or references that have a diachronic aspect to them, but are present-oriented. Finally, we used the [ecology] tag to mark references that include attention to the biological or ecological linkages that connect social economies of dried fish to their environments. The remaining three tags are linked to two objectives of DFM. A key part of the project’s motivation is the intention to generate applied insights in nutrition and health in particular. These are areas, of course, linked to wellbeing and social and economic relations. Our project recognizes that existing and potential governance systems and policy regimes facilitate or constrain action, and therefore must be an important area of attention.

Topic	Search terms	Initial thematic tags	Revised thematic tags	Thematic cluster
economics	demand value chains		value chain / microeconomics	

	economics	economy		value chains, economy, and labour
	---	labour		
health	health	health		food and nutrition security and health
	food security	nutrition and food security		
	nutrition			
social relations	gender	gender		culture and social relations
	women			
	wellbeing	well-being	culture, social relations & well- being	
	---	culture		
	ethnography	---		
	---	food and cooking		
	---	history / change		
	---	political ecology		
technical interventions	technology	Technical	food engineering	food science
			food chemistry and microbiology	
			food safety	
	---	policy and governance		ecology, policy, and governance ⁶
---	ecology			

TABLE 4. Evolution of our thematic categorization. Original Google Scholar search terms were matched to thematic tags, two of which were combined (“well-being” and “culture” into “culture, social relations & well-being”) and one of which was split (“technical” into “food engineering”, “food chemistry and microbiology”, and “food safety”) during the course of analysis. In the final stage, these tags were grouped into five thematic clusters, as discussed below.

⁶ ##THEME: ecology, policy, and governance” represents a kind of “other” category, catching applied themes that are significant to the general fisheries literature (natural resource management, fisheries policy, governance) but that are not yet explicit in dried fish literature.

COMPUTER-ASSISTED ANALYSIS

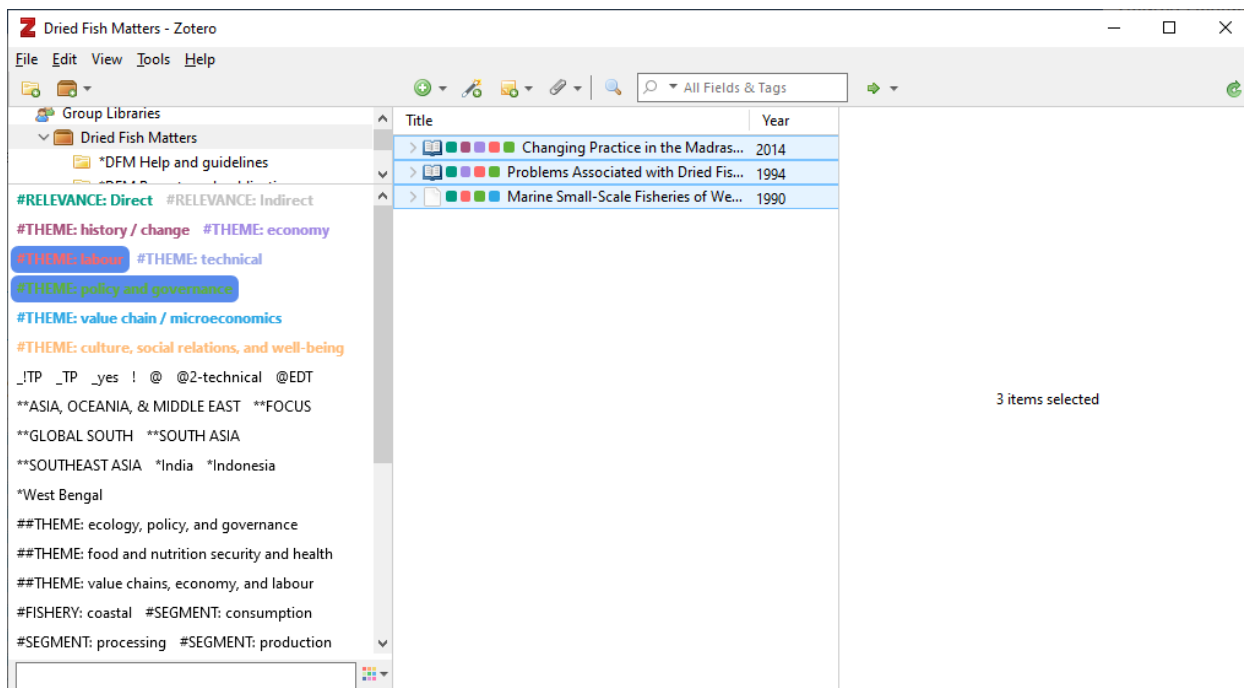


FIGURE 4. Manual analysis of tags using the Zotero desktop client.

It is clear that although we could have completed a review of the dried fish literature without the assistance of Google Scholar or Zotero, in the absence of those tools we would almost certainly have sought to conduct a review of significantly reduced scale, or used familiar sources to produce the background section for an original research work. We might have looked in greater depth at available resources for selected qualitative evidence to support our definitional claims, without attempting to provide a quantitative survey of our literature census. In working with Zotero, we found that the easiest place to begin our study was with a quantitative review, as the software allowed us to retrieve quick counts of publications tagged with a given keyword or set of keywords. While these counts could be generated quickly within the desktop program interface (fig. 4), we were quickly drawn to the possibility of automating this process through the Zotero's Application Programming Interface (API), a command-based interface designed to be accessed over the web by computer scripts. The Zotero API (Stillman 2021) allows users to build tools that query or update a Zotero library, essentially replicating the functionality of the desktop client by assigning URL "commands" to each operation that would be performed interactively with a keyboard and mouse. An open-source library exists to facilitate simple queries using scripts in the Python programming language, which can be as simple as three or four lines of code (Hügel et al. 2022). For example, the following code prints the titles of the first three items in the Dried Fish Matters Zotero library ("2183860" is the identifier of our group library):

```
from pyzotero import zotero
zot = zotero.Zotero('2183860', 'group')
items = zot.top(limit=3)
```

```
for item in items:  
    print(item['data']['title'])
```

We found the Zotero API useful in generating a spreadsheet item counts by tag, the largest volume of which reflected geographic regions. During our tagging process we had applied tags representing over 100 countries and regions, which we subsequently wished to aggregate for analytic purposes into various types of groupings – continents, Global North/South, “focus region” (South and Southeast Asia), and so on. For each of these geographic regions, we expected to see significant differences in coverage of specific themes or product types within the dried fish literature. The figures showing these correlations would have taken many hours to compile manually, and – most importantly – we would have been discouraged from repeating the analysis following further interaction with the Zotero library. By automating this process, we only needed to list the tags we wished to combine and correlate once; this allowed us to review and present preliminary findings, then use those findings to refine our searching and tagging efforts, leading to an iterative analysis. We were also able to adjust our geographic categories, for example by shifting Middle East resources alongside Asia instead of Africa.

As our work progressed, we developed our Python scripts into a command-line utility that not only allowed us to query the library items by tag, but also to: create and apply “cluster” tags aggregating geographic regions, themes, or product types, making these available for analysis in the desktop client; compare the tags in the library against a list of tags in a working list; create a spreadsheet listing the number of articles in the library published in each journal represented in the database; or print a bibliography containing a list of tags under each item (Thrift 2022). A command that generated bar graphs out of tag union queries was spun off into a web-based script, designed to take lists of tags and graph options through an online form and return basic quantitative measures about the distribution of references matching the user query in the form of an html table or bar graph (Thrift 2021b).

VISUALIZATION AND KNOWLEDGE CONSTRUCTION

This visualization utility allowed us to play around interactively with a variety of search parameters to generate the series of graphs that were ultimately included in our literature review, and to update figures as the library data evolved. It also shaped our conceptualization of the dried fish literature, not simply by drawing our attention to visually apparent differences in thematic coverage between different regions, but also by encouraging us to *construct* visual difference between different themes, through repeated experimentation with graph and data parameters to achieve the most legible graphs. For example, Figure 5 represents an initial attempt to map significant themes by major geographic region, which we subsequently revised through modification of our analysis parameters. Each bar in the graph represents the number of publications in the dried fish literature that was tagged as belonging to a particular theme, with bars colour-coded by theme and clustered by region. Although this graph accurately conveys the quantitative data present in our database, it is not particularly legible: the “technical” bar for Asia, Oceania, and Middle East dwarfs all the other themes, causing the scale of the graph to make distinctions in smaller clusters less intelligible; the overall distribution of references within each cluster is evident, but it is difficult for us to perceive the total number of

publications for each theme; and the large number of colour-coded categories makes it difficult to read the graph without looking back and forth repeatedly between bars and legend.

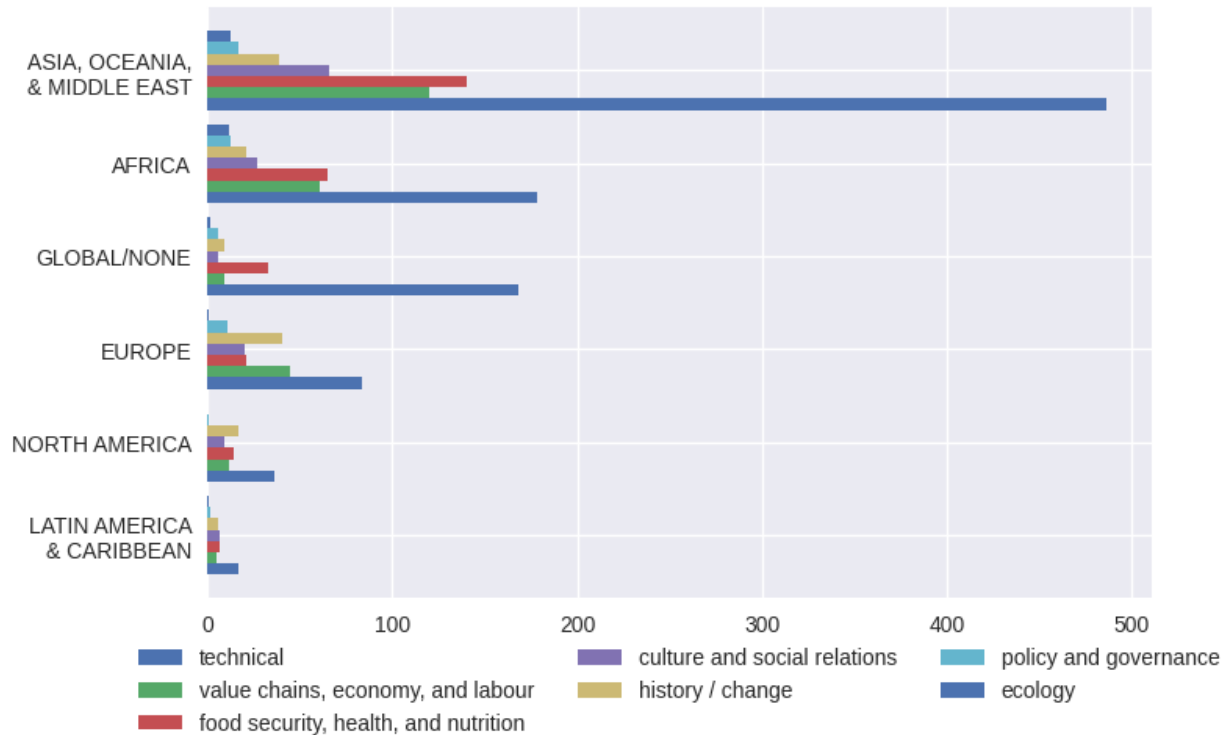


FIGURE 5. Original bar graph showing the distribution of themes by continent in the Zotero library, by count of items matching each combination of theme/continent tag. As tags are not mutually exclusive, some sources are represented more than once. All items in the library were tagged with at least one thematic tag and at least one geographic tag. In an editorial comment on this graph in a draft of our literature review article, one of the co-authors wrote: *This figure isn't too useful in its current format since there are too many categories. There are two options to deal with this: (1) conflate the regional categories (e.g., "Asia & Oceania", "Americas", "Africa", "Europe") or (2) generate a selective set of bar graphs for individual themes. We can probably exclude the themes like "ecology" that have almost no coverage anyway.*

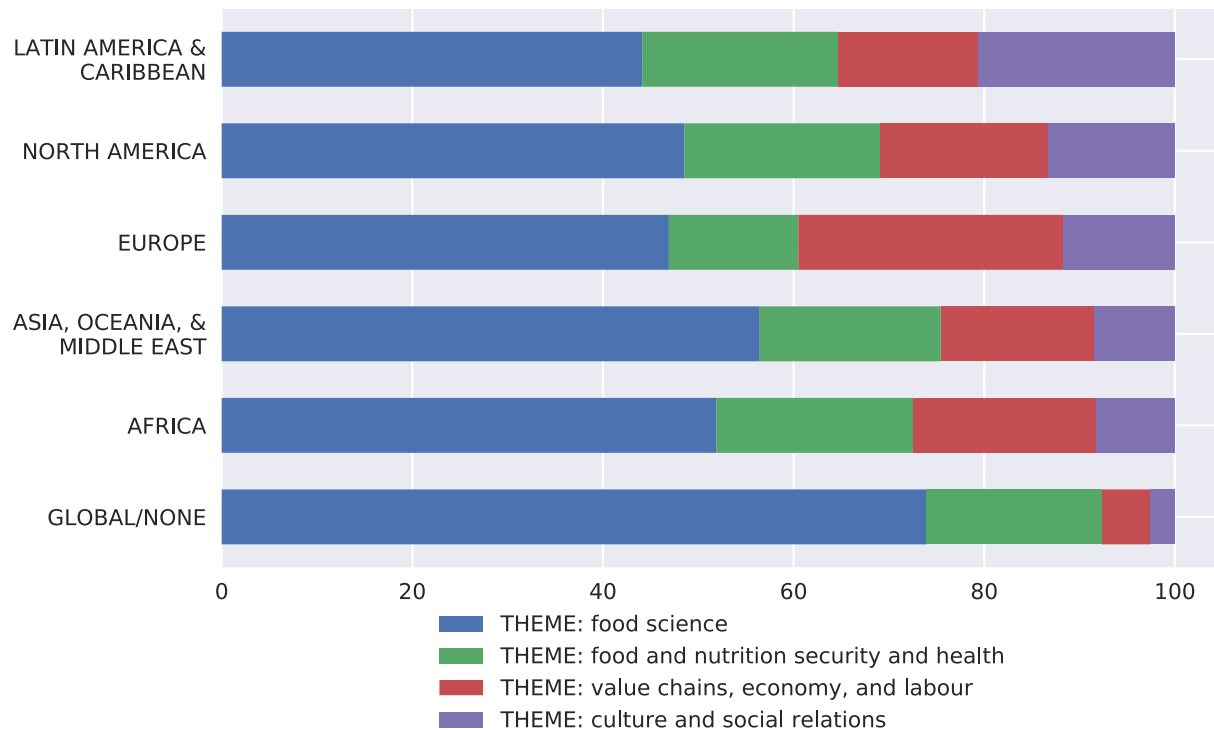


FIGURE 6. Themes by continent: updated graph.

A second version of this graph (figure 6), also unused in our final review, displays the same data as stacked bars – but here we have only four thematic categories instead of seven, representing clusters of themes. An intermediary version of this graph presented the same clustered bars as percentage of the overall literature; now we have percentage-stacked each group, reducing the content of the graph to show only the relative proportion of texts for each region that address a particular theme. The grouping of themes was not an analytic imperative, but one that was originally driven by visualization concerns; yet it ended up shaping our final analysis, in leading us to assert that the dried fish literature is organized by a particular set of value concerns corresponding to these legible clusters – *dried fish as food*, *dried fish as commodity*, and *dried fish as cultural heritage*. Similarly, we created cluster categories of fish products in order to simplify representation (fig. 7), opting to represent the categories dried fish, salted fish, fermented fish, smoked fish, fish sauce, and “other”. These categories absorbed the further five categories of fish crackers, fish paste, fish powder, pickled fish, and derivative products. The choice to limit our representation of the literature themes to proportions of sources by continent speaks to the reductive demands of our work, whereby our efforts to capture the essential features of the literature topography were driven by the visual medium available through the graphing tool.

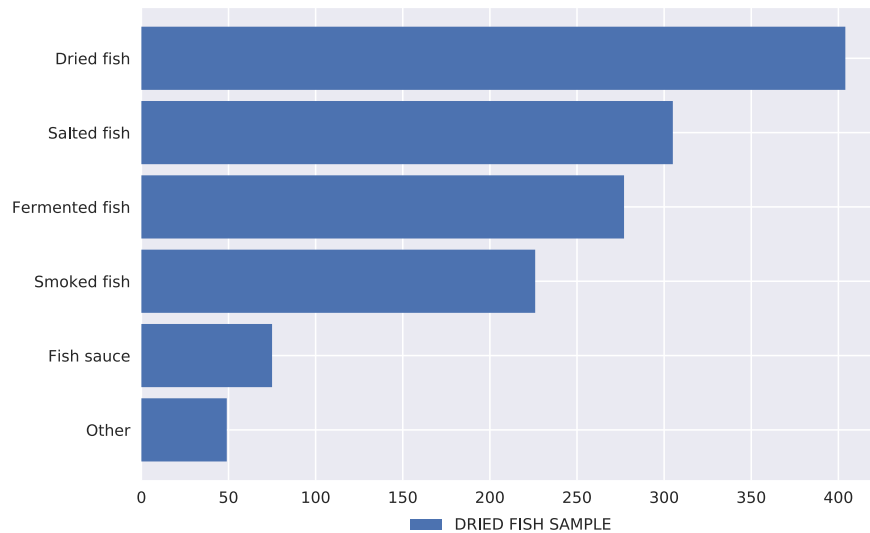


FIGURE 7. Dried fish product types. All items in the Zotero library were tagged with zero or more product types.

Insofar as the quantitative section of our review was structured as a discussion of these graphs, our interpretive commentary also downplayed or excluded points related to graphs that were cut from the review due to poor visual effectiveness, while highlighting points of strong visual interest. For example, in our discussion of Figure 8, we observe “proportionately greater coverage of the Global South than is apparent in the general fisheries literature sample ... suggesting that the literature on dried fish mirrors global geography of fish production and consumption more closely than the general fisheries literature”. We also suggest that “the literature on dried fish in Asia accounts for nearly half of tagged references ... Africa and Europe also each have substantial literatures, while references on the Americas are much smaller proportionately”. Significantly, each of these points is made as an interpretation of the accompanying graph, rather than the graph being constructed as an illustration of these claims.

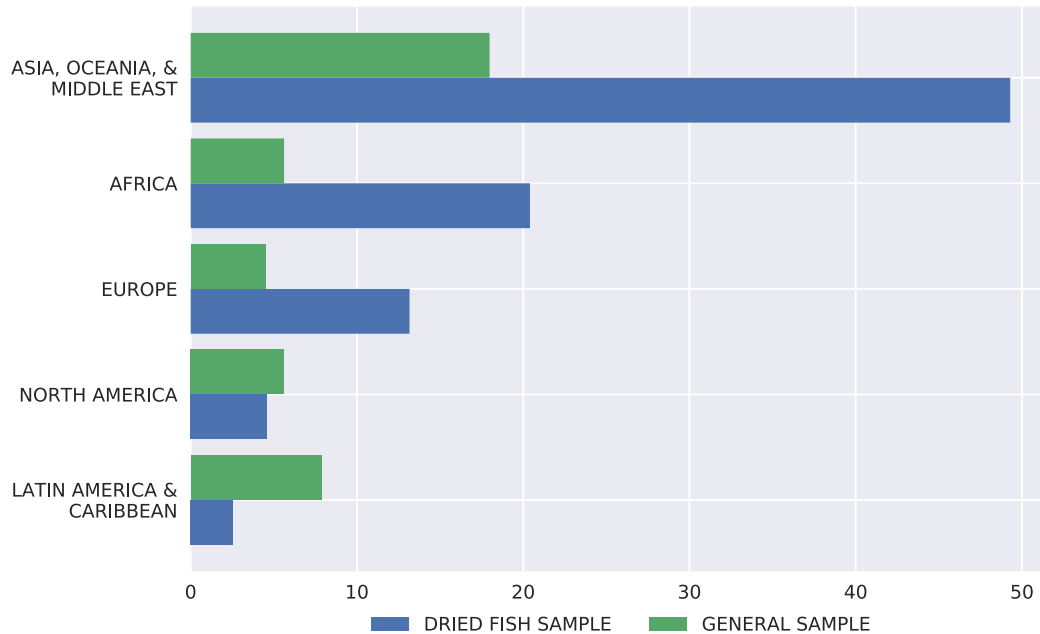


FIGURE 8. Geographic comparison of the dried fish literature to a “general fisheries” sample, by percentage of items in each sample tagged with each continent-level geographic tag.

Our sequence of tagging, quantitative analysis, and finally (qualitative) discussion was established in the interest of convenience, as quantitative results could be achieved soonest through a more superficial triage and tagging of the literature. But this organization resulted in the qualitative analysis being structured by the quantitative review. A review of a smaller, less statistically valid literature sample might have discouraged attempts at quantification, instead facilitating an inductive approach to pulling out key findings, comparing and contrasting publications, and discussing the implications of significant ideas that emerged from close reading of the texts in that sample. By contrast, our quantitative study guided a more structured and deductive approach, whereby we read through sets of publications representing themes and approaches that we considered *a priori* “representative” within the literature, and focused more on refining our definition of the *dried fish literature* by seeking out representative publications, outliers, and sub-thematic distinctions.

REFERENCES

- Belton, Ben, Derek Johnson, Eric Thrift, Jonah Olsen, Mostafa Hossain, and Shakuntala Thilsted. Forthcoming. "Dried Fish at the Intersection of Food Science, Economy, and Culture: A Global Review." *Fish and Fisheries*.
- Corporation for Digital Scholarship. 2020. "Zotero Item Types and Fields." https://www.zotero.org/support/kb/item_types_and_fields.
- . 2022. "Zotero: Your Personal Research Assistant." 2022. <https://www.zotero.org/>.
- Dried Fish Matters. 2021. "Global Dried Fish Literature Review Database." University of Manitoba. <https://doi.org/10.34990/FK2/UEECDM>.
- Food and Agriculture Organization of the United Nations. 2022. "AGRIS Homepage." 2022. <https://www.fao.org/agris/>.
- Google. n.d. "Google Scholar Help." Accessed January 28, 2022. <https://scholar.google.com/intl/en/scholar/inclusion.html>.
- Halevi, Gali, Henk Moed, and Judit Bar-Ilan. 2017. "Suitability of Google Scholar as a Source of Scientific Information and as a Source of Data for Scientific Evaluation—Review of the Literature." *Journal of Informetrics* 11 (3): 823–34. <https://doi.org/10.1016/j.joi.2017.06.005>.
- Hügel, Stephan, Peter Gerdes, Sky Bristol, and Pablo Orduña. 2022. "Pyzotero 1.5.1 Documentation." 2022. <https://pyzotero.readthedocs.io/en/latest/>.
- Ivey, Camille, and Janet Crum. 2018. "Choosing the Right Citation Management Tool: Endnote, Mendeley, Refworks, or Zotero." *Journal of the Medical Library Association : JMLA* 106 (3): 399–403. <https://doi.org/10.5195/jmla.2018.468>.
- Johnson, Derek, Benjamin Champion, Ratana Chuenpagdee, Milena Schreiber, Jessie Varquez, and Ansen Ward. Forthcoming. "Small Fish Processing." In *Small and Low-Cost Fish for Food Security and Nutrition: A Food Systems Approach*. FAO Technical Papers. Food & Agriculture Organization of the United Nations.
- Johnson, Derek, Jonah Olsen, Eric Thrift, and Ben Belton. 2019. "Comparing Economies of Dried Fish in Asia: What the Literature on Six Countries Tells Us." Conference paper presented at the MARE People & the Sea X Conference: learning from the past, imagining the future, University of Amsterdam, June 26.
- Johnson, Derek, and Eric Thrift. 2020. "Using Zotero to Structure a Global Literature Review on Dried Fish." Presented at the TBTI Transdisciplinarity course, Online, October 8.
- Martín-Martín, Alberto, Enrique Orduna-Malea, Mike Thelwall, and Emilio Delgado López-Cózar. 2018. "Google Scholar, Web of Science, and Scopus: A Systematic Comparison of Citations in 252 Subject Categories." *Journal of Informetrics* 12 (4): 1160–77. <https://doi.org/10.1016/j.joi.2018.09.002>.
- Martín-Martín, Alberto, Mike Thelwall, Enrique Orduna-Malea, and Emilio Delgado López-Cózar. 2021. "Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: A Multidisciplinary Comparison of Coverage via Citations." *Scientometrics* 126 (1): 871–906. <https://doi.org/10.1007/s11192-020-03690-4>.
- Stillman, Dan. 2021. "Zotero Web API Documentation." Zotero Documentation. November 5, 2021. https://www.zotero.org/support/dev/web_api/v3/basics.

- Thrift, Eric. 2021a. "Online Tools for Collaboration in a Time of COVID-19." Conference paper presented at the MARE 2021 People and the Sea Conference, Online, June 30.
- . 2021b. *Zotero Tags Graphing Tool*. Python. Dried Fish Matters.
- . 2022. *Zotero-Meta-Analysis-Toolkit* (version 1.0). Python. Dried Fish Matters.
- Thrift, Eric, Ben Belton, and Derek Johnson. 2020. "Dried Fish in Global Perspective: Themes and Knowledge Gaps from a Multidisciplinary Literature Review." Webinar, December 10.
- Winslow, Rachel Rains, Sarah L. Skripsky, and Savannah L. Kelly. 2016. "Not Just for Citations: Assessing Zotero While Reassessing Research." In *Information Literacy: Research and Collaboration across Disciplines*, edited by Barbara J. D'Angelo, Sandra Jamieson, Barry Maid, and Janice R. Walker, 287–304. The WAC Clearinghouse; University Press of Colorado. <https://doi.org/10.37514/PER-B.2016.0834.2.14>.