

A systematic survey of online trade: trade in Saiga Antelope horn on Russian-language websites

DAVID L. ROBERTS, KATYA MUN and E. J. MILNER-GULLAND

SUPPLEMENTARY MATERIAL 1 A systematic online search strategy for wildlife trade products.

Wildlife trade, both legal and illegal, is increasingly moving onto the internet. As a result, strategies and protocols need to be developed to help provide an understanding of the nature of this online trade. The Supplementary Material presented here provides a general strategy for conducting a systematic survey of online trade in wildlife products. Here, the trade in saiga antelope *Saiga tatarica* horn is used as an example, however the strategy is intended to be generalisable to all forms of online wildlife trade. This strategy draws on systematic evidence review protocols, in particular the Cochrane Handbook (<http://handbook.cochrane.org/>).

Before starting a survey it is important to understand and identify who the stakeholders are. For the purposes of transparency and agreement with the outputs it may be beneficial to work at all stages with stakeholders (e.g. a focus group).

Identifying the question and objectives of a survey

As with any systematic search it is important to identify the scope of the search and particularly the question/s being addressed. The question should be well-formulated so that it allows for the identification of the main concept/s. If the question is broad, it may be beneficial to list specific objective/s and also inclusion criteria to help hone the development of the search. Below is a suggested way of structuring the objective/s that covers the behaviour of the trade, what exactly is being traded, on what internet platform, in/from which country and the period over which the trade is taking place.

To determine

- the <trade behaviour (e.g. extent, interaction)>
- of trade in <population e.g. taxon (e.g. saiga), production origin (e.g. wild v captive), geographical origin (e.g. Russia), product (e.g. horn)>
- over <platform (e.g. internet, social media, eBay)>
- in <country of trade (e.g. Russia), possibly including origin of trade (e.g. internal only, both internal and trade into)>
- during <period of trade (e.g. last 12 months)>

It is critical that due consideration is given to the question and objective/s as this will guide the rest of the search and prevent wasting time collecting data that is not relevant. It is particularly important if this search strategy is being used as a metric such as to determine whether demand reduction has been effective.

Ethics & governance considerations

As a note of caution, as the purpose is to conduct a systematic survey of online trade, trade involves human subjects, therefore ethics and governance issues need to be considered

depending on the intent of the survey and its outcomes. The area of cyber-ethics is rapidly evolving and therefore we do not provide a review of the considerations here. However, by its very nature studying online trade is covert, as it is difficult to gain prior informed consent; although it is not necessarily deceptive. The data collected in such surveys consist of elements that vary in the degree to which they are considered personal data. As a result provisions need to be made regarding data handling and storage. Other aspects to consider include, but are not limited to, copyright and contract law regarding data access.

Starting a search

Where you start your survey can potentially bias what you find. The two main search methods are via the use of search terms or by browsing a specific website using the hierarchical structure.

Search terms – The most common method is through the use of search terms such as “saiga horn” + “for sale”. Using such search terms allows you to search specific websites (e.g. eBay), as well as using generic search engines (e.g. google). The main problem is that you will only find what you search for and this can impact on your perception. A good example of this is if we search for elephant “ivory” on eBay, besides the issue that “ivory” is a colour as well as a material, knowledgeable sellers rarely use this word, instead they use code terms to disguise the fact they are selling ivory that is banned on eBay. As a result, a search using the term “ivory” would find little if any elephant ivory for sale and conclude that there is little elephant ivory being sold through eBay, when in fact there is a considerable amount depending on the code words used. Alternatively, using a generic search engine may not be able to identify trade on certain sites, particularly closed forums on social media sites (e.g. WeChat). Therefore when conducting a systematic search it is important to recognise the limitations of the methods one is using.

Browsing – On specific websites it may be possible to use a browse function through subcategories to identify trade, however this again will bias any results, for example on eBay UK ivory can be found for sale under oriental antiques while in France it tends to be in the subcategory of ethnographic antiques.

Here we will only consider the use of search terms, rather than a browser-based search strategy.

Search terms

With the wildlife trade it may be difficult to identify keywords or phrases owing to the range of species involved, the fact that a product may undergo several stages of transformation in its pure state, what the end products are and the range of languages involved. However, providing the original question and objectives have been well-formulated this should help reduce the number of search terms.

Two strategies can be employed to identify the keywords and phrases to be used, namely (1) a focus group of experts in the field, and/or (2) snowball sampling, that is by using initial search terms in a pilot study to identify associated words. However, before starting consideration needs to be given to the language in which the search is being conducted.

To identify the search terms the language/s must first be defined; it is important to note that this is an online survey so it is how the language is written rather than spoken. In the case of

the trade in saiga horn in Russia we may be interested in Cyrillic and Roman. In certain cases, English may be used as a common language (lingua franca) between individuals whose native language is different. However, in some cases a language, such as English, may be used as a conduit to allow discussion and identify the terms in, for example, Russian, depending on the common language amongst the experts.

Logic grid

The systematic investigation of wildlife trade over the internet can be considered analogous to a systematic evidence review. In systematic evidence reviews a Population, Intervention, Comparison intervention, and Outcome measures (PICO) logic grid can be used to help visualise the elements and keywords. Here the PICO logic grid can be modified to look specifically at the components of the objective statement outlined above. While the internet platform (unless one is specifically interested in a type of platform e.g. auction sites or forums), country or countries in which the search will focus (although note this could be more international if the researcher is looking at trade in or out, i.e. international/cross border, of a specific country as well as trade within i.e. domestic, and by the very nature that some internet platforms are international), and the duration are dealt with elsewhere in terms of the search, here we are interested only in those components that will form the search terms, specifically the population and the trade behaviour.

Population – Considering the population, a logic grid may consist of Taxa, Origin, Pure product transformation and Finished product (TOPF) (including trademarked product names) (Supplementary Table 1); it may be beneficial to also add Language and Manufacturer. However, this will vary depending on the question and objective/s, for example we may only be interested in the pure rather than the finished product and therefore the finished product column may be blank. Here is an example for the trade in Saiga horn (Supplementary Table 1).

SUPPLEMENTARY TABLE 1 TOPF (Taxa, Origin, Pure product transformation and Finished product) logic grid for Saiga Horn trade.

Taxa	Origin	Pure product transformation	Finished product
<u>Saiga</u>	<u>Wild</u>	Horn	Cooling water
Saiga	Wild	Horns	Handle
Saigi		Shavings	Knife
Saigachi	<u>Farmed</u>		Pills
Antelope	Captive	Por	Skull cap
	Farmed	Pora	Trophy
<i>Saiga tatarica</i>			
<i>Saiga t. tatarica</i>			Нож

Saiga t. mongolica

Пчак

Рукоятка

Сайгак (nominative, masculine)

Сайгака (genitive, masculine)

Сайги (genitive, feminine)

Сайгачьи (possessive adj.)

Антилопа (nominative)

Антилопы (genitive)

Trade as a term – While we are interested in say saiga horn, we are specifically interested in the trade, rather than any discussion on saiga horn. Search terms are therefore needed that describe an item being traded/offered for sale, such as words associated with sale (potentially wanted) and also the currency used (Supplementary Table 2). This is particularly important when using search engines that are generic or websites that discuss the product, rather than specific websites that are focussed purely on the sale of a product (e.g. eBay). The same linguistic considerations apply in relation to the terms associated with trade as described previously.

SUPPLEMENTARY TABLE 2 Logic grid for trade terms.

Terms associated with trade	Currency
Buy	Ruble
For Sale	Rouble
Price	Рубль
Sale	₽
Want	
Куплю	
Продам	
Продаю	

From these grids we can then identify those words that are expected to result in the greatest number of positive hits (we may not want to search for all) and also create search phrases that incorporate TOPF and trade/currency terms. Again a focus group may be employed to finalise the search terms. However, before starting a search or even when designing a search

it is important to have an understanding of how searches work on specific websites and what advanced search functions allow one to do. For example in systematic evidence reviews using databases such as ISI Web of Science (WoS) and SciVerse SCOPUS, it is standard practice to use Boolean operators and wildcards; this may or may not be possible to do depending on the website being used. As a result, it is not only important to have a good understanding of how a website's search function works but also any variation in the spelling of a term, particularly if it is not possible to use wildcards.

Checking for term redundancy – As part of a pilot study or at the end of a full study it may be useful to check for redundancy in the terms used to inform future studies. An obvious example that should be identified when initially selecting the search terms is the use of hierarchical terms such as “saiga” and “saiga horn”; where “saiga horn” will return a subset of items that would be found using the term “saiga”. However, as part of a pilot study it may be interesting to determine if using a specific term such as “saiga horn” misses important items when compared with “saiga”. In some cases the redundancy may be less obvious where two, seemingly related, terms return a high proportion of items that are found in both searches (i.e. both terms have a high association). It may be more economic in terms of time to only use one of these search terms rather than both. Related to this in some cases, a search term may result in such a low number of relevant ‘hits’ or that there is a low ratio of relevant ‘hits’ to the number of items one needs to search. Again some may feel use of such terms is uneconomical in terms of time.

Conducting a search

At all stages, but particularly when conducting a search, it is important to reduce bias in terms of the items found and at the same time ensure reproducibility. During the construction of the question, objective/s and identification of search terms, some consideration will have been given to the platform on which the search is going to take place. However, it is also important to consider how the data, often from items, will be recorded. For example, what does one do if an item appears in repeated weeks, or across different websites or an item offers several individuals for sale? This has important implications in terms of understanding the nature of the trade as well as independence of data points when it comes to analysis of the data.

As mentioned previously, typically during systematic evidence reviews, specific databases such as ISI Web of Science (WoS) and SciVerse SCOPUS are used. These have specific search protocols in relation to the use of Boolean operators and wildcards, which may not be possible when searching websites. Further, while WoS and SCOPUS are obvious starting points for a search of the literature, the starting point for a survey of wildlife trade is less clear, particularly when it is not targeted to a specific website, therefore this, along with the search terms as discussed, can influence the results. One possible option is to use ranked lists of website usage (e.g. <http://www.alex.com/topsites> although it may vary in reliability depending on the country) which may help identify target websites (e.g. the highest ranked social media website for a country) or an appropriate search engine for more generic searches. However it should be noted that these ranking websites could be somewhat biased and in some cases trading platforms may not fall into the website’s rankings.

Surveys using a generic search engine

If a specific website is the target of the study (e.g. eBay), then the search protocol will largely be dictated by the search facility of the website; often they have advanced search facilities. We will therefore only consider searches using a generic search engine such as google.com. Assuming a search engine has been selected (it should be noted that two different search engines are unlikely to give you the same result), the question and objective/s will determine what advance functions will be required, such as limiting the search to the last 12 months, in a specific language and/or country, and exact word or phrase (search terms).

It should be noted that the list of items a search engine generates might change, firstly through marketing methods whereby companies can increase their position in a search result. For example, beyond a website containing words that appropriate search terms, regularly updating the content, manipulating/adding the metadata within the website that describes the page, adding links, and using alt tags, can increase a website's position in a google search. Secondly, related to the first point of the example, by repeatedly visiting a website based on a search term can increase its position. This latter point needs to be taken into consideration by those undertaking the survey. This is often termed as a filter bubble; this results in searches becoming tailored to the user. Here the website algorithm 'guesses' what information a user would like to see. To do this it uses information about the user including location, websites they have clicked on in the past (click-behavior) and their search history. As a result the user is effectively 'trapped' in their own cultural or ideological bubble. In the case of conducting a systematic survey of online trade, a researcher using their 'normal' computer to conduct a search may well be receiving targeted recommendations for them, rather than a 'clear' set of results. As such, it may be worth using a new profile in the web browser, deleting cookies or finding other ways of avoiding the filter bubble. A more advanced option, that allows a researcher to conduct repeated searches using the same specifications, increasing replicability and standardizing data retrieval, is to use a search engine's API.

A search can result in thousands of hits, not all will be relevant and it is likely the majority of those that are relevant will be towards the top of the list, however it may be worth checking in a pilot study. As a result it is likely that the results may be limited to the first X number of samples, for example the first 200. Limiting the results as described has the added benefit over limiting by the number of results a research can get through, as one person could be faster than another. However, if the search engine lists the most relevant results at the top, then statistics based on the proportion of relevant to non-relevant items cannot be extrapolated to the total number of hits. Alternatively, a stopping rule can be applied, such as when less than 1 in 10 items is deemed relevant. This can also be applied once say the first 200 items have been considered. Either way the search protocol needs to be applied consistently across all search terms.

Screening for inclusion

In a systematic evidence review following a search of the literature, the literature is filtered first by the Title, then by Abstract and finally by the Full text. At each stage statistics are collected and the reasons for exclusion included. For a trade survey it is likely that the criteria will only apply to the title (i.e. what the search generates), followed by directly inspecting the post. However, the item titles as generated by a search engine may be of limited use as a filter as they may contain insufficient information, therefore all items generated may need to be checked.

To ensure that the inclusion criteria are effective and may be reproduced by others, potentially in the future, not only is it good practice to clearly document the processes involved in the search but also verify the results are reproducible. One method is for the person conducting the survey to sit down with a volunteer and identify a number of items (e.g. 10-15% of the total with a minimum of 100) and identify whether they should or should not be included. The level of agreement can be checked and reported using Kappa analysis; items where there are differences should be discussed.

Snowball sampling websites

Once the searches have been completed through a general search engine (e.g. google.com) it will hopefully result in hits for wildlife trade products. These will include items for sale on auction sites, classified ads, online shops and social media. These sites themselves can then be searched using the same strategy as for the general search engine, although this will depend on the functionality of the search option on the website.

Additional searches

Beyond search engines and searching the resulting websites (i.e. snowball sampling websites) it may be beneficial in some cases to search other sites that did not arise during the general search. These may include websites that rank highly in terms of usage (e.g. see <http://www.alex.com/topsites>), or from a search of previous studies that indicate sites they have monitored.