

Online Appendix:

What Information Do Citizens Want?: Evidence from One Million Information Requests in Mexico

2018, *World Development*

Daniel Berliner

Assistant Professor of Political Science and Public Policy, Department of Government
London School of Economics and Political Science

D.Berliner@lse.ac.uk

Benjamin E. Bagozzi

Assistant Professor of Political Science and International Relations
University of Delaware

bagozzib@udel.edu

Brian Palmer-Rubin

Assistant Professor, Department of Political Science
Marquette University

brian.palmer-rubin@marquette.edu

Contents:

1. Additional Details on Methods
2. Examination and Description of Topics
3. Maps of Topic Distributions
4. Unstemmed Topwords in Spanish

Additional Details on Methods

Sample and Preprocessing

In the main paper, we briefly discuss our sample of Mexican access to information (ATI) requests (2003-2015) and the preprocessing steps that are used to prepare this corpus for analysis. The current subsection of the appendix provides a more extensive discussion of these aspects of the project. Recall that each and every Mexican federal ATI request has been made publicly available in CSV format on the Mexican Government's Sistema Infomex website (INAI, 2015) for the period June 2003 until August 2015 (at the time of analysis). We downloaded each of these CSV files for the 06/2003-08/2015 period,¹ which contain individuals' yearly ATI requests themselves, links to any attachments associated with these requests, text of agency responses to each request, and a host of other relevant meta data related to the timing of the request and the timing and nature of the Mexican government's response. For this sample frame, Mexico's ATI requests varied in frequency by year, ranging from 37,732 in 2004 to a peak of 148,545 in the year 2013, with an average number of requests per year of 103,830.²

The focus of the present paper is on the ATI request texts themselves for the 2013-2015 period. These request texts appear within the “DESCRIPCIÓN SOLICITUD” field of the downloaded request meta-data and correspond to each requestor's own open-ended description of the specific information that they are requesting from the federal government of Mexico. We hence sought to retain all primary request texts that were provided within this field for use in our primary analysis. However, we *do* exclude from this request sample all information requests for confidential personal data (i.e., “Datos Personales”), as these personal data request texts (i) are covered under a different legal regime, and (ii) have been redacted within the CSV files described above for privacy reasons. Above and beyond the remaining “DESCRIPCIÓN SOLICITUD” request texts (i.e., all non-personal information requests) that we retain in our sample, two additional fields within the CSV meta-data also include highly relevant text pertaining to the content of information requests: attachments and “otros datos”. We choose to add these two additional fields into our primary set of request texts, and describe how we do so immediately below.

First, recall that while the vast majority of requestors described the nature of their requests within this field (in Spanish), a smaller subset either left this field blank, or in 128,695 cases (i.e., roughly 10% of our total sample of requests) included a portion (or more commonly, all) of their request as a separate file attachment. As these attachments are highly relevant to our analysis, we implemented a webscraping routine to download each and every attachment based upon its archived web-link provided within the “ARCHIVOADJUNTOSOLICITUD” field of the aforementioned CSV files. These request-attachments came in a wide variety of file-types, but were primarily submitted as PDF, Microsoft Word (e.g., .docx), Microsoft Excel (e.g., .xls), CSV, or image (e.g., .png, .tiff) files. After downloading, we converted all attachment files and file types to machine readable text using a variety of conversion programs specific to each file

¹ At the time of analysis, the latest publicly available requests corresponded to August 31st, 2015.

² I.e., when omitting 2003 and 2015 requests from our calculations of these summary statistics given that we only observe requests for part of these years.

type and (in the case of the PDF attachments) optical character recognition (OCR) software.³ We then incorporated these converted texts into our primary set of information request texts.

Second, and beyond the attachment files described immediately above, an additional field within the request meta-data is labeled “OTROSDATOS” and is also directly related to request text content. Based upon our readings of both the request process and a large sample of request meta-data, “OTROSDATOS,” though often left blank, is nevertheless used by a small number of requestors to provide additional information clarifying or extending the information that was requested within the primary DESCRIPCIÓN SOLICITUD field, and/or within their attached materials. Like our attachment materials, this information is highly relevant to the content of individual’s information requests, and we therefore merged all text found within the “OTROSDATOS” field to our request texts where applicable.⁴

The combined, “DESCRIPCIÓN SOLICITUD”, attachment, and “OTROSDATOS” set of request texts described above comprises our final sample of text for analysis. However, before we can proceed to analyze this corpus with our intended topic modeling techniques, we need to apply a range of additional “preprocessing” steps to our corpus to prepare it for topic modeling analysis, as well as to ensure that our topic models yield meaningful, representative, and interpretable information from this corpus. We describe our specific preprocessing steps in detail immediately below.

First, while the attachment request texts that we merged into our sample together represent a large proportion of all Mexican Federal ATI requests made during this period (i.e., 10%)—and appear to be highly relevant to our efforts to ascertain the nature of the information that Mexican citizens seek from their government—the addition of these attachments to our final corpus created one moderate challenge. Specifically, a very small number of the attachments that we converted to machine readable text and added to our final set of request texts were exceptionally long (e.g., hundreds or thousands of pages) and appeared to correspond to large PDF (or Excel) files that included a small amount of relevant request information followed by a larger combination of example documents or spreadsheets with lists of names, municipalities, and/or medical supplies. Given the latter features, much of the content contained within these exceptionally long information requests likely arises from a different data generating process than that which is generating our primary request texts of interest. Thus, for the final combined request entries⁵ with more than 1,000 character strings,⁶ we truncated these documents from the thousandth string onwards. In this fashion, we are able to retain any relevant information contained within the front-matter of these exceptionally long request texts while also discarding a majority of these requests’ more auxiliary, “list based” content. Given that only 0.02% of our

³ A very small number of attachments—fewer than 0.1% of all attachments—were corrupted or missing based upon their provided links and/or uploaded documents hence were not converted.

⁴ For the latter, we took care to omit any “OTROSDATOS” entries that simply read “OTROS DATOS” or “ninguno.”

⁵ That is, for any one of our final, combined “DESCRIPCIÓN SOLICITUD”, attachment, and “OTROSDATOS” request text documents.

⁶ By character strings, we mean all words, unique number sequences, or special character sequences separated by spaces.

final documents contained more than 1,000 character strings, this decision affected a very small number of our final sample's actual request texts.

For all remaining request texts, we applied a set of common natural language processing (NLP) techniques so as to convert our final corpus of ATI requests into a collection of individual “documents” that are compatible with the unsupervised topic modeling techniques used in our main paper, as well as to fix common typographical errors (i.e., typos) that appeared across our request texts. Regarding the latter, it bears mentioning that because requestors (of varied educational background and writing abilities) were typically hand typing their requests into an online interface, typos and grammatical errors were a relatively common phenomenon. In light of this, we first endeavored to convert all grave accents appearing within our corpus to acute accents as the former are not used in Spanish. Second, some requestors did not include *any* relevant accents within their requests, and we hence sought to standardize the use of accents for *some* commonly appearing words that are known to systematically have accents in Spanish.⁷ Third, we further removed any and all nongraphical characters that appeared throughout the corpus, which corresponded to a small number of uncommon symbols arising from the OCR-ing of our PDF attachments. Fourth, we removed all websites, based upon character strings beginning with either “www” or “http,” which most frequently occurred within our converted attachment texts.

We next processed each Spanish request text to remove all punctuation, numbers (in numeric, verbal, and roman-numeral forms), names of months, and stopwords. The removal of these character-sets is standard preprocessing for our anticipated topic model techniques (Roberts et al., 2014; Bagozzi and Berliner, 2017; Bagozzi 2015) and helps to eliminate extremely common, but highly uninformative, character strings. We then also removed (i) roughly 65 commonly used bigram or trigram phrases that were commonly used by requestors to indicate that the information request is not provided in the text field, but rather is attached (e.g., “se adjunta archivo”), (ii) instances where individuals chose to leave the request field blank by filling it in with arbitrary text,⁸ and (iii) several additional problematic character strings that are prone to arise within extreme large corpora of unstructured texts, namely floating letters (e.g., “a”, “q”, “s”). Next, our preprocessing routine leveraged a TF-IDF representation of our corpus to omit sparse terms that do not occur in at least 0.1% of the documents in our corpus. This is a common means of eliminating extremely rare words that offer little insight into the shared topics across one's text corpus (Quinn et al., 2010; Grimmer, 2012; Rice, 2012). Finally, and as is also standard in probabilistic topic modeling (e.g., Bagozzi and Berliner, 2017; Bagozzi 2015), all remaining words were converted to lower case, stemmed, and re-structured into a document-term-corpus (using term frequency as opposed to TF-IDF) before analysis. Altogether, the preprocessing steps discussed above created a corpus with 1,023,267 processed request documents and 6,696 total unique words.

⁷ We could not standardize all words, as many Spanish words have two different meanings with and without accents. Nevertheless, words ending in, e.g., “ion,” “pon,” “don,” “ton,” “zon,” “eria,” “isimo,” as well as all relevant names of Mexican states, and the word “Mexico” itself, were standardized where appropriate (while also taking care to adjust any cases of exception)..

⁸ For example, entries that just repeated the request-field's name (“DESCRIPCIÓN SOLICITUD”) or included a series of “x's” (e.g., “xxxxxxxxxx”).

Modeling Approach

As discussed in the main paper, we analyzed this corpus with a widely used unsupervised topic model known as Latent Dirichlet Allocation (LDA; Blei et al., 2003). LDA is a probabilistic topic model that is primarily designed for finding groupings of words that characterize latent dimensions of texts. As such, it is commonly used in social science applications for the discovery of the underlying, “latent” issues or themes that are systematically discussed within and across a given corpus of text documents (e.g., Bagozzi and Schrodt, 2012; Rice, 2012; Barberá et al., 2014; Bagozzi, 2015). In social science applications such as these, the documents themselves are assumed to be a mixture of multiple, overlapping latent topics, each with a characteristic set of words. LDA estimates these topics via a three-level hierarchical Bayesian model that treats each document as containing a mixture of underlying topics, where the topics themselves are specified as an infinite mixture over a corresponding latent set of topic probabilities (Blei et al., 2003, 993). In this manner, each identified topic can be characterized by the vector of words that LDA assigns as having the highest probability of association with that topic. These topic-word associations can then be used to (i) interpret each topic for its substantive meaning and/or (ii) to classify one's documents in terms of the topic(s) most likely to have generated each document, as based upon the observed distribution of words within each topic.

More technically, assume that LDA is applied to a collection (i.e., corpus) of M documents wherein each document d is a sequence of N words (w_{dn}). In such a case, $\mathbf{w} = \{w_1, w_2 \dots w_N\}$, and LDA assumes that each document in M is comprised of a mixture of underlying topics, which it then seeks to estimate from the observed distributions of words across and within documents. The proportion of document d corresponding to a given topic k is represented as π_{dk} , such that $\pi = \{\pi_{d1}, \pi_{d2}, \dots, \pi_{dk}\}$ across all topics. LDA then proceeds in two general estimation steps. First, document proportions are drawn from a k dimensional Dirichlet prior, $\pi_{dk} \sim \text{Dirichlet}(\alpha)$, with shape parameter α . Second, and for each document d with word positions $n \in \{1, \dots, N\}$, LDA draws a word's topic, $\tau_{dn} \sim \text{Multinomial}(1, \pi_d)$, and then conditionally draws a word (w_{dn}) from $w_{dn} \sim \text{Multinomial}(1, \eta_k)$, given τ_{dn} . One then estimates an approximation of the (log) likelihood for the entire document collection by taking the product of each document \mathbf{w} 's corresponding marginal probability $p(\mathbf{w}_d)$. More extensive treatments of the technical details to LDA estimation, and related approaches, can be found in Blei et al. (2003), Liénou et al. (2010), and Grimmer and Stewart (2013).

As the above estimation framework implies, the LDA technique fully discards the orderings of words within sentences and documents, and instead looks to more simply identify “clusters” of words that frequently co-occur with one another across the documents in one's corpus. As such, LDA is inherently a “bag-of-words” approach to automated text analysis, in that it uncovers thematic meanings from texts based upon observed correlations between words, rather than based upon the syntax of these words. While bag-of-words assumptions accordingly entail a loss of important textual information, they have nevertheless been shown to yield coherent substantive insights across a wide variety of political science applications (Quinn et al. 2010,

Grimmer 2010, Hopkins and King 2010; Roberts et al. 2014, Bagozzi 2015). Indeed, as Hopkins and King (2010) note, while this bag-of-words approach “seems counterintuitive at first, since it is easy to write text whose meaning is lost when word order is discarded (e.g., ‘I hate Clinton. I love Obama’) [...] empirically, most text sources make the same point in enough different ways that representing the needed information abstractly is usually sufficient” (pg. 232). Given that our Mexican ATI request corpus has well over one million documents, we are very much able to leverage the latter features described by Hopkins and King for our application. This makes bag-of-words approaches such as LDA an ideal tool for the comprehensive content analysis of the full sample of 1,023,267 Mexican ATI request texts, 2003-2015.

That being said, there are a large number of additional bag-of-words approaches beyond LDA, and some further discussion of our choice of LDA over these alternate approaches is warranted. One simpler alternative to LDA for identifying representative words within a corpus of documents is the TF-IDF statistic. TF-IDF weights the frequency of a word’s appearance in a given document by that word’s inverse document frequency, where the latter quantity can be defined as the logged total number of documents in a corpus divided by the number of documents containing that word. As such, TF-IDF allows one to identify the words in a corpus that are most unique to a given document or set of documents, which can be used, in turn, to identify and represent clusters of documents in a similar fashion to LDA. That being said, achieving the latter tasks with TF-IDF requires a larger number of user-specified assumptions and post hoc adjustments, relative to the LDA approach used below. In addition, as Blei et al. (2003) note, “[TF-IDF] also provides a relatively small amount of reduction in description length and reveals little in the way of inter- or intradocument statistical structure” (pg. 994). Given the large and complex nature of our corpus, these TF-IDF limitations are especially acute for our application, leading us to avoid the use of TF-IDF as our primary analysis tool. Note, however, that we do leverage TF-IDF weightings at the intermediate preprocessing step of sparse term removal—as described in our preprocessing steps above—before converting our remaining retained terms back to document term frequencies for use in our LDA models.

Returning to other plausible bag-of-word models of our corpus, one important set of distinctions can be made between our use of LDA on the one hand and the alternatives of either (i) “single membership” topic models such as K-Means or (ii) “mixed membership models” such as the topic model proposed by Quinn et al. (2010) on the other. Single membership models obligate that each document arise from only a single topic, whereas mixed membership models such as the model developed by Quinn et al. (2010) allow for mixtures of topics to occur at any given point in time (or some other unit of aggregation), but nevertheless require that each document arise from a single topic. By contrast, LDA is a mixed membership model that assumes that *each document* is a mixture of underlying topics, where this mixture can vary across different documents depending on the words contained within each document. We direct interested readers to Grimmer and Stewart (2013) for a more detailed discussion and comparison of these approaches, as well as of mixed and single membership models more generally.

In our case, we believe that the LDA model best fits the texts and data generating process for our Mexican ATI corpus. While many ATI request texts in Mexico very likely pertain to only a

single issue area,⁹ many others *do* encompass multiple themes, especially once one has taken into account the larger amounts of text that is commonly included within request attachments. In such cases, mixed membership models, such as LDA, allow the researcher to recover the ‘best of both worlds,’ in effect by allowing us to estimate mixtures of topics within those documents where mixtures of themes are actually present, while also (if performing correctly) estimating some documents as arising disproportionately from a single topic when single membership is in fact the case.¹⁰ Indeed, given that we find that roughly 100,000 of our final ATI request “documents” are estimated as having 80% or more of their text arising from a single topic, it would appear that LDA is able to handle and identify single topic requests. Moreover, and owing partly to this flexibility, extant research has shown LDA to be highly useful and effective in recovering coherent social science topics from documents whose lengths are far shorter than the documents analyzed here, including the first 3-4 lead lines of international newswire reports (Bagozzi and Schrodtt 2012) and postings on the social media website Twitter (Barberá et al., 2014).

The above points notwithstanding, it is also important to recognize that our corpus of ATI requests contains a large amount of structural information, including the day (and year/month) of a given request, the location (i.e., municipality and zip code) provided by each requestor, and the target agency of the information request. In employing LDA, we are largely ignoring this auxiliary estimation during our estimation of topics. This is in contrast to a number of more recently developed topic models—including those of Quinn et al. (2010) and Grimmer (2010), as well as the Structural Topic Model (STM; Roberts et al. 2014) – each of which, to varying degrees, conditions the estimation of topics upon this auxiliary structural information. We choose not to employ this structural approach in our current analysis for the following interrelated reasons. First and foremost, our objective at present is to establish a baseline set of results for topics of ATI requests in Mexico for future research to build upon. In these respects, LDA helps us to ensure that these baseline topical findings (i) have made a minimal number of assumptions¹¹ and (ii) can be readily extended into the greatest number of future directions.¹² Secondly, our partitioning of structural information from our estimation of actual topics also helps us to ensure that the subsequent temporal and topical variation that we do examine in the main paper is not hinging entirely on our choices to include some, but not all, of the wide and varied auxiliary meta-data contained within the CSV files discussed above. Nevertheless, we firmly believe that future analyses of specific features of the Mexican ATI corpus analyzed here will find the incorporation of auxiliary structural information into the topic model estimation stage to be extremely useful and beneficial.

⁹ Indeed, we classify our request texts as arising from their most “dominant” topic along these lines in the main paper.

¹⁰ This is in contrast to single membership models, which would “force” each document to arise from a single topic.

¹¹ I.e., in terms of what auxiliary information (i.e., fields of our CSV meta-data, including municipality vs. zip code, year/month/day, agency, etc.) matters and should be used to condition topics on, and which should not.

¹² As LDA remains by far the most familiar topic model across the sciences and industry, as well as across different software platforms, in addition to being the common starting point for the vast majority of the topic models that have been developed in more recent years, including the STM.

Topic Number Selection

Returning to our discussion of the LDA model and our primary ATI request-text corpus, we next provide further background on our selection of an optimal number of topics to be estimated by our final LDA model. Here, recall that for LDA, as well as for the many other topic models mentioned above, researchers must explicitly choose the number of topics, k , to be estimated. In this regard, we agree with others (e.g., Roberts et al., 2014) that there is no “right” choice for one’s k , and that this decision should primarily rest on substantive expertise and the degree of granularity of interest. Even so, researchers have used a number of summary quantities to inform their choice of topic number for LDA models, where LDA-derived measures of topical density have been employed as a means of identifying (i) the readability level of documents within the arena of educational development (Qumsiyeh and Ng, 2011), (ii) the number of issues discussed within international climate change negotiations (Bagozzi, 2015), and (iii) the representativeness of a document sample with respect to the themes found within a broader document collection (Baille et al., 2009). The most common metric used in determining an optimal number of underlying topics identified by LDA across a set of documents is the perplexity measure, which can be defined for a given test set S_{test} of M documents (d) as

$$\text{perplexity}(S_{test}) = \exp \left(- \frac{\sum_{d=1}^M \log(p(\mathbf{w}_d))}{\sum_{d=1}^M N_d} \right)$$

where a lower perplexity score indicates better generalization performance (Blei et al., 2003; Liénou et al., 2010, 30). Less frequently, the marginal likelihood for one’s LDA model, which is monotonically increasing in one’s perplexity statistic, is used as an auxiliary measure of model fit and topic number (e.g., Grün and Hornik, 2011, 28).

However, others have cautioned against the (over)use of model fit diagnostics for the selection of one’s topic number within probabilistic topic model applications (e.g., Roberts et al. 2014). Instead, more qualitative assessments of classified documents and representative topwords, obtained from topic models using different pre-assigned values for k , is often offered as an alternative approach. Bagozzi and Berliner (2017), for example, estimate topic models of $k = \{10, 15, 20\}$ and qualitatively examine the topwords and representative documents obtained from each before choosing a topic model that uses $k = 15$ in their analysis of U.S. State Department Human Rights Reports. As discussed in more detail in the main paper and further below, there are a range of topwords that we can extract and compare in these regards, including the words with the highest probability of association with each topic, as well the words *most highly associated* with each topic according to frequency exclusivity scoring metrics, which ensures that our reported topwords correspond to the words that are both most frequently assigned to a given topic and most exclusive in their assignment to that topic. Moreover, for our specific application, we can examine these highest probability or most highly associated words in either their original (stemmed) Spanish form, in their unstemmed Spanish form,¹³ or in English translations of their unstemmed Spanish forms. While we preliminarily examine *each* type of

¹³ Where, for a given stemmed word, we simply query the entire corpus for its most common unstemmed counterpart.

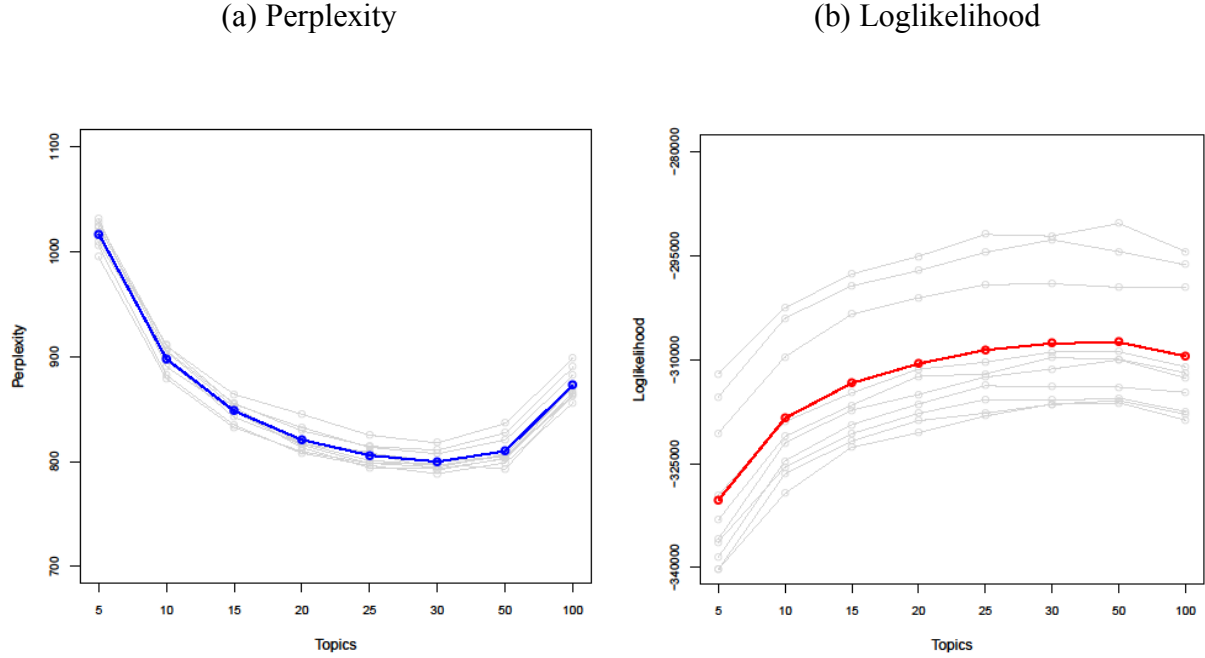
word vector (i.e., most highly associated and highest probability; stemmed-Spanish, unstemmed Spanish and English) in evaluating the different k 's discussed below, we follow others (e.g., Roberts et al. 2014; Bagozzi and Berliner 2017) in focusing our discussions on the *most highly associated* topword vectors below.

In the paragraphs above, we outlined a series of (i) model fit diagnostics and (ii) qualitative assessment approaches that researchers have previously used to select optimal numbers of topics within social science topic modeling applications. To identify the optimal topic number for our own Mexico ATI request analysis, we jointly employed both sets of approaches. Below, we first discuss the model diagnostic evaluations that were implemented to guide our final choice of k , and then present a set of qualitative-assessments that were used to complement, and validate, our aforementioned model-diagnostic methods for topic number selection.

To implement a model-diagnostic driven selection routine for k , we employ each of the model fit statistics that were introduced above alongside a cross-validation sampling routine. Specifically, we use a ten-fold cross-validation approach to identify the optimal number of topics underlying our corpus. Here, we first draw a random sample of 10,000 information requests from our final preprocessed text corpus and then randomly divide this sample into ten folds of training and test data.¹⁴ For each set of training data, we next estimate a series of LDA models where the number of topics, k , is sequentially set to $k = \{5, 10, 15, 20, 25, 30, 50, 100\}$. We then use each resultant LDA model's output to initialize a validation (LDA) model using our corresponding test sample. With these models in hand, we calculate each validation model's perplexity and loglikelihood measures at each k . Altogether this approach—which is comparable to that used by Grün and Hornik (2011) and Bagozzi (2015)—yields ten sets of perplexity measures, and ten sets of loglikelihood measures, for each k evaluated (i.e., $k = 5, 10, 15, 20, 25, 30, 50, 100$). These individual values are plotted in grey Figure A.1, and the mean values of our perplexity and loglikelihood values are then calculated and overlaid in blue and red, respectively.

¹⁴ Note that for each fold of test data, the remaining data in the corpus was used as that fold's training data.

Figure A.1: Model Selection



For Figure A.1, recall that one prefers lower perplexity measures but higher loglikelihood values. Hence, Figures A.1a-A.1b suggest that an optimal number of topics for our corpus rests somewhere in the $k = 20$ to $k = 50$ range, since this general range clearly yields the lowest and highest values for our perplexity and log-likelihood measures. In order to guard against overfitting, we follow extant applications of these metrics (Barberá et al., 2014; Bagozzi, 2015) to identify a topic number at the lower end of this scale, $k = 20$, as the optimal topic number in this case. Indeed, one can observe that the gains to be had in loglikelihood and especially perplexity by increasing k beyond 20 are marginal to negative, especially relative to the observed improvements in each metric when one increases k from, e.g., 5 to 10. Furthermore, while increasing k beyond 20 does offer some improvement in perplexity and loglikelihood, our choice of the more conservative value of $k = 20$ is consistent with the “one standard error rule” (Hastie et al., 2009),¹⁵ in that our perplexity value at $k = 20$ is the smallest k that still falls within one standard error of our lowest overall perplexity value (i.e., at $k = 30$).

While the above model selection routines provide strong support for a 20-topic LDA model, they also offer some support across the broader $k = 20$ to $k = 50$ range. We therefore estimated a subset of these LDA models when using our full text corpus, in an effort to validate our topic-number findings more qualitatively. Specifically, we estimated a set of LDA models on our full

¹⁵ Which has been employed as justification for selection of k in past LDA applications (Barberá et al., 2014).

preprocessed request text corpus using Gibbs sampling while setting our LDA model's α parameter to the frequently assigned value of 0.1 (e.g., Yano et al. 2009, Bagozzi and Schrodtt 2012; Bagozzi 2015, Priva and Austerweil 2015, Ahonen 2016), which has been noted elsewhere as in keeping with “common topic-modeling practice” (Ahonen 2016: 61).

In our application, this choice for α is also based heavily upon our own qualitative understanding, and theoretical expectations, of the content of the information requests in the Mexico context. As discussed in the main paper, as well as above and below, our corpus of Mexican ATI requests generally encompasses citizen, journalist, and business-initiated requests made to specific Mexican federal agencies for fairly specific topical information. Hence, while we do expect *some* mixing of topics within the information requests in our Mexican ATI corpus—especially within our longer requests (e.g., those containing attachments)—a substantial number of our requests are fairly short in length (e.g., 1-10 sentences), and it is unlikely these requests—or our corpus’ requests more generally—simultaneously contain a mixture of *all* potential issues of concern that are commonly raised within Mexico’s federal ATI request system. For example, while it is reasonable to expect that a typical information request may touch on 1-3 separate topics related to, e.g., distributive programs *and* personal information; it is highly unlikely that the average information request in our corpus would to jointly request information on distributive programs, crime, public procurement, medical supplies, budgets and spending, the environment, and education—to name just a few potential topics. Our close reading of 1,600 actual ATI requests (discussed further below) supports these conjectures.

The above rationale helped to inform our choice of the hyperparameter α for our final LDA models. A *relatively* low α value (such as 0.1) in this case minimizes the likelihood that our LDA model treats our request documents as containing mixtures of all—or close to all— topics, which would be the case for higher values of α . As such, our choice of a low α ensures that our LDA model will assume that our documents are each generally composed of a small number of topics, rather than of most-to-all topics. Given that—as mentioned above—our request texts are typically fairly short and topically-focused, we believe this is a desirable assumption, as it is unlikely that *a majority* of our requests contain a mixture of all—or even most—of the latent topics that exist across *all* request documents in our corpus. At the same time, this choice of $\alpha = 0.1$ is also not as extreme as to force all documents to follow this process, by, e.g., forcing only a single topic to correspond to each document, which thereby ensures that the assumptions that we are imposing in this case are not overly atypical or extreme. Indeed, as mentioned above, this precise value of α has been utilized in past LDA applications to political texts (e.g., Bagozzi and Schrodtt 2012; Bagozzi 2015, Priva and Austerweil 2015, Ahonen 2016) and is validated as an appropriate α value for our specific application within an auxiliary cross validation routine discussed further below.

After assigning our α for these full sample comparison models, we then incrementally set the number of topics to be estimated within these LDA models to $k = \{10, 20, 30\}$. This range was chosen based upon extant applications of LDA and related topic models to social science texts, where extant scholars have reported optimal topic models of 15 (Bagozzi and Berliner, 2017), 20 (Bagozzi and Schrodtt, 2012), 25 (Bagozzi, 2015), and 30 (Barberá et al., 2014) for corpora of similar content to our own. After estimating each of these LDA models, we then extracted the stemmed and unstemmed Spanish versions of each identified topic’s top twenty words according

to both the words with highest probability of assignment to each topic, and the most highly associated words with each topic according to frequency and exclusivity.

Based on evaluation of the topwords, and close reading of sampled requests associated with each topic, we determined that a 20-topic model offered the best balance of detail, parsimony, and interpretability. In the 10-topic model, many topics were difficult to interpret as they clearly combined several different common types and themes of request. The 30-topic model, on the other hand, offered largely redundant additional detail over the 20-topic model, as many of the additional topics simply further split common themes into a larger number of smaller topics; such as yielding yet additional topics pertaining to procurement, government employees, and medical supplies.

Thus, our model fit analyses, our qualitative assessments of topics, and extant research together provide strong support for the choice of a 20-topic LDA model within our primary analysis. The interpretation of the resultant topics from this 20-topic model (both further below and in the main paper) reinforces this topic number choice. We next turn to a more detailed presentation of our primary 20-topic model and its results.

Topics

As discussed in the main paper, we estimated our final LDA model on our full corpus of 1,023,267 preprocessed documents, as opposed to the 10,000 document sample used for parameter selection above. For this LDA model, we use a Gibbs sampling routine within the `topicmodels` package in R (Grün and Hornik 2011),¹⁶ set $k = 20$, and again assign our LDA model's α parameter to the value of 0.1. Regarding the latter choice, and as argued above, we believe theoretically that our ATI requests will typically contain a mixture of small numbers of (similar) topics, rather than mixtures of most-to-all topics. This helps to justify our choice of $\alpha = 0.1$, in that a low (but still widely accepted) value on this hyperparameter ensures that our LDA model treats our documents as generally being composed of a small number (i.e., subset) of topics. Secondly, we also verify this be a reasonably appropriate α value for this application via an auxiliary cross-validation routine. Specifically, we repeat the cross-validation approach that was used to identify our optimal k above, in this case by first randomly drawing a wholly distinct set of 10,000 from those used in our earlier cross-validation exercise. We then divide this new sample into 10 folds, and evaluate a set of training and test-LDA models on each fold, and repeat this process for α parameter values of 0.01, 0.1, 0.5, 1, and 5. In examining the corresponding perplexity values (averaged across all 10 folds) for each potential α parameter choice, we find that our chosen α parameter value of 0.1 yields by far the best performing perplexity measure (of 835.6), and thus LDA model, followed second by the α value of 0.01.

Altogether, the final LDA model discussed here accordingly identifies the 20 topics that best characterize our ATI request texts across the entire 2003-2015 period. Each of these topics represents an underlying word distribution where every word in our corpus is given a posterior probability of assignment to that topic. For these word vectors, and consistent with extant research (Roberts et al., 2014; Bagozzi and Berliner, 2017), we then again derive the words *most*

¹⁶ Here using single chain of 3,000 iterations (treating the first 2,000 iterations as burn in).

highly associated with each topic according to frequency exclusivity scoring metrics, which ensures that our reported topwords correspond to the words that are both most frequently assigned to a given topic and most exclusive in their assignment to that topic. It is up to us to substantively interpret the meaning of each topic based upon these word assignments.

In order to label our topics appropriately, we first examined the most frequent and exclusive topwords (in both Spanish and English) for each topic. We then classified each of the request-text documents in our corpus according to its single most associated topic. To classify our documents in this manner we employ our LDA model's posterior topic distribution parameters, $\hat{\pi}_{dk}$, which, for each document (d) provide a probability estimate ($\hat{\pi}$) of that document's association with each of our twenty topics (k). Using these sets of k probabilities, we identified the topic with the highest probability of association for each document, and classify our documents accordingly. While this approach sacrifices a degree of information, it is standard in the literature (Grün and Hornik, 2011; Bagozzi and Schrodt, 2012; Bagozzi, 2015).

In addition, summary statistics indicate that the typical request in our sample does exhibit at least one topic whose probability of assignment can be seen as “dominant.” Indeed, the median and mean posterior topic distribution probabilities for each document's maximum $\hat{\pi}_{dk}$ value are 0.47 and 0.48, respectively, suggesting that the documents in our corpus typically have one topic that nearly outweighs all others in terms of probability of association.

Using these dominant topic classifications, we then conducted a close reading of 1,200 individual requests, comprising samples of sixty requests predominantly associated with each topic. Among requests that were associated more strongly with a given topic than with any other, we sampled twenty requests each at high, medium, and low topic probabilities for that topic, where those reflected probability ranges above 0.8, between 0.5 and 0.7, and between 0.3 and 0.5, respectively. Based on these example requests as well as the topwords for each topic, we assigned labels to each of the twenty topics. These output quantities are presented and discussed in more detail within our main paper and our “topic appendix”.

Examination and Description of Topics

This appendix provides a detailed description of the twenty topics that were identified by the primary unsupervised topic modeling approach presented in the main paper. We discuss our process for assigning descriptive labels to each topic and our qualitative appraisal of the potential of each topic for generating social accountability. We based this process on the following types of information for each topic: close reading of 80 randomly sampled information requests from each topic (totaling 1600 requests in all), lists of each topic's 20 "topwords" and "FREX topwords," the distribution of requests across agencies, and the distribution of requests across states.

In the text below, we summarize this analysis and provide four representative examples of information requests in each topic. In these examples we censor individual names, business names, reference codes for registrations or court cases, or any other information contained in requests that may compromise the anonymity of requesters.¹⁷ Otherwise, these example requests are presented as is, including any spelling or grammatical errors or inconsistent capitalization or punctuation. We also provide brief summary translations of each example request.

In our analysis of individual information requests in each topic, we read 80 requests, selected through four random samples of 20 requests each. Half of these requests were sampled from the set of requests that most exemplified each topic. That is, for each topic, we drew two samples of 20 requests each from the set of requests with topic probabilities above 0.8 for the given topic. However, we also wanted to ensure that our interpretations were not driven entirely by these "high-certainty" requests, and also sought to review requests with lower topic probabilities, but still a clear "majority" or "plurality" probability of being associated with the given topic. We thus examined two additional samples of requests for each topic: 20 from the set of requests with topic probability between 0.5 and 0.7 for the given topic, and 20 from the set of requests with topic probability between 0.3 and 0.5 for the given topic. By analyzing requests with different probabilities of topic assignment, we were able to evaluate the degree to which the topics consistently sorted requests.

Our substantive interpretations of each topic, discussed in this appendix, are based on close readings of these samples of requests, the topwords for each topic, and additional information pertaining to the distribution of each topic's requests across the federal agencies they were sent to, and across the states they were filed from. Combining these multiple sources of information, and our substantive familiarity with politics, civil society, and access to information in Mexico, yielded clear interpretations for nearly all twenty topics. We make clear also those topics where we are less confident in our interpretations. We present these descriptions in order from the topic accounting for the greatest share of requests to that accounting for the smallest share. We retain the topic numbers originally output by the topic modeling procedure, but these numbers convey no meaningful information except as unique identifiers.

¹⁷ While we use only data made publicly available by the government of Mexico, and requesters are warned by the INFOMEX system that their requests will become publicly available, some requests nonetheless include names or email addresses.

In the topic descriptions below, we report the top ten topwords as well as the top ten FREX topwords. Recall that our topic evaluations included assessments of the top twenty topwords and top frex words associated with each topic. In the topic descriptions below, we report a selection of these quantities—the top ten topwords and top ten FREX words—to illustrate our approach. The topwords correspond to the words that are assigned the highest probability of association with a given topic based upon the posterior word distribution for each topic obtained from our model. The FREX (frequent and exclusive) topwords are the words that are both (i) most frequently assigned to a given topic (based again upon each topic’s posterior word distribution) and (ii) the most exclusive in their assignment to that topic. While the topwords are useful for characterizing each topic, the FREX words are often more informative as they filter out words that are common to many requests and topics, such as references to the information request system, niceties, or other very common terms (e.g., document, program, case, information), and thus “provide more semantically intuitive representations of topics” (Roberts et al. 2014, 1068).

Importantly, for both the topwords and FREX topwords reported below, we present *destemmed* versions of our model’s recovered (top) word-stems. Recall that prior to topic modeling, all words in our corpus were reduced to their word-stem (base or root) for comparability. In order to enhance interpretability of our model’s resultant topword and FREX topword vectors, we have hence destemmed all topwords for presentational purposes below by identifying the most frequently occurring complete word-match to each word-stem (within our full, raw corpus of requests), and replacing that word-stem with this more complete counterpart. As multiple complete words typically have the same word-stem, this process provides a representative, but non-definitive, example of each word-stem’s corresponding destemmed version.

The top five agencies reported for each topic are based on the reported proportions of requests in that topic that are submitted to each agency rather than the proportion of all requests to that agency that are in that topic. This focus on within-topic variation allows us to diagnose the type of issues that the topic addresses without results being skewed by the volume of requests sent to each agency. We followed a similar procedure in determining the geographical distribution of requests (unreported).

In discussing the potential for accountability for each topic, we infer the potential requesters of information for each topic and the likely uses. These uses may be accountability generating—such as monitoring government provision of goods or services—or may serve private purchases—such as finding out information useful for applying for government benefits, winning a government contract, or checking on a regulatory evaluation. We also reference the potential “action cycles” (Fung et al. 2007, 51-54) through which the information requested could be used in generating accountability, such as through media campaigns, lobbying, or electoral campaigns. This process is necessarily speculative and many topics include a mix of requests that appear to be accountability generating and private.

Topic 3: Land and the Environment

This topic primarily concerns environmental issues, environmental impact reports, construction, land tenure, land use, and planning. This is reflected both in the topwords, and in the frequency of requests directed to ministries dealing with the environment and natural resources. This is the largest topic, accounting for 8.6 percent of all requests. However, it is also the most geographically diffuse topic, with the lowest Herfindahl index of concentration across states. This reflects the prevalence of local interests in requests for this type of information. This topic's share of all requests has been relatively stable, with no clear time trend.

Many individual requests in this topic ask for environmental impact reports for specific construction or infrastructure projects, other official planning documents, property titles, information on boundaries of protected areas, or information on environmental damage. However, the topic model also captures some unrelated requests, particularly when they include names of specific neighborhoods or municipalities, or terms specific to land.

Likely requesters are private citizens affected by development projects, local civil society or environmental groups, as well as journalists, lawyers, or firms involved in land or environmental disputes. This topic has a high potential for accountability-seeking uses. Information may be deployed to make claims on the state regarding such public issues as environmental enforcement and infrastructure, and private claims such as processing of land titles, and permits. However, some requests may also come from developers or construction companies themselves, seeking information for commercial purposes.

Comparative Measures

- Percent of requests: 8.6% (1/20, largest topic)
- Herfindahl index across agencies: 0.069 (7/20, 7th most concentrated topic)
- Herfindahl index across states: 0.129 (20/20, least concentrated topic)

Topwords:

- municipio, ubicado, zona, proyecto, agua, solicitado, construcciones, san, ambiental, federal
- municipality, located, zone, project, water, request, buildings (construction projects), san, environmental, federal

FREX words:

- ejido, predio, terrenos, sueldo, forestal, superficie, tramite, semarnat, minera, ambiental
- ejido, property (land), plots of land, wage, forest, surface area, bureaucratic procedure, SEMARNAT (Federal Ministry of the Environment), mining, environmental

Top Five Agencies (with proportions of requests in this topic)

SECRETARÍA DE MEDIO AMBIENTE Y RECURSOS NATURALES	0.197
COMISIÓN NACIONAL DEL AGUA	0.117
SECRETARÍA DE COMUNICACIONES Y TRANSPORTES	0.096
REGISTRO AGRARIO NACIONAL	0.034
PROCURADURÍA FEDERAL DE PROTECCIÓN AL AMBIENTE	0.032

Examples

January 12, 2004: SECRETARÍA DE ECONOMÍA

- Listado de proyectos de construcción de carreteras autopistas y libramientos carreteros que se tiene proyectado construir este año. Incluir entidad federativa y longitud de la carretera.
- (List of planned highway construction projects, including ministry and length of road.)

November 27, 2006: SECRETARÍA DE COMUNICACIONES Y TRANSPORTES

- los planos oficiales COZ/RP-007 COZ/RP-011 y COZ/RP-014 denominados Delimitación y Determinación de la Franja Costera del Municipio de Cozumel Quintana Roo georeferenciados
- (Official boundary plans for protected coastal area in Cozumel, Quintana Roo.)

March 25, 2013: PROCURADURÍA FEDERAL DE PROTECCIÓN AL AMBIENTE

- Información sobre contingencias ambientales derrames de hidrocarburos deterioros ambientales o denuncias de vecinos referentes al tema ambiental. El sitio esta ubicado el el lote D y lote E (entre av. San Lorenzo y Paseo de servidumbre) en el Parque Industrial FINSA Puebla Autopista México - Puebla Km. 117 Puebla Pue. 72710
- (Information on environmental damage, oil spills, or environmental complaints for a specific location in an industrial park in Puebla.)

April 8, 2013: SECRETARÍA DE MEDIO AMBIENTE Y RECURSOS NATURALES

- El expediente completo incluyendo anexos alcances información adicional manifestación de impacto ambiental resolutivos y correcciones sobre la obra PLANTA MUNICIPAL DE TRATAMIENTO DE AGUAS RESIDUALES DE LA CIUDAD DE HERMOSILLO con la clave de proyecto de SEMARNAT [project code].
- (Environmental impact report for a wastewater treatment plant in Hermosillo.)

Topic 6: Individual Needs

Requests in this topic are notable for their orientation towards individual needs. Most are questions about how to access specific government services or how to complete specific procedures. This is reflected in the topwords, which do not share a substantive theme but rather capture verbs such as “to do,” “to know,” and “would like,” and other words frequent to simple requests focused on the needs of the individual requester. This is the second largest topic, accounting for 8.5 percent of all requests. This is also the second most diffuse topic, both in terms of the states from which requests originate, and the agencies to which they are sent. The low geographic concentration reflects the “mundane” nature of these requests and their association with the needs of ordinary citizens all over the country. This topic also does not capture any particular substantive policy areas that are the domain of any one agency, but rather a common theme of requests that many different agencies often receive.

This topic’s share of all requests has trended downward over time, possibly reflecting a reduced need for individuals to seek these types of information via official request, instead being able to use government websites via the internet, or other sources.

Many individual requests in this topic ask for how to access government services such as social benefits, pensions, education, or subsidized loans; or how to complete official procedures or fulfill requirements. Some of these requests ask *where* to go or inquire to fulfill these needs, while others ask *what* specific requirements, forms, or information is necessary. However, the topic model also captures some unrelated requests, particularly questions about government activities asked in broad and simple language. These are likely grouped together due to similar language as the relatively simple requests relating to individual needs.

Likely requesters are ordinary citizens seeking assistance or guidance for their own private purposes, as well as inexperienced information requesters asking questions in non-specific terms. Some appear to be commercial entities as well, but similarly seeking assistance with official procedures. This topic has very low potential for accountability-seeking purposes, as requests typically seek information useful for making private decisions or problem-solving, rather than information that may relate to state performance.

Comparative Measures

- Percent of requests: 8.5% (2/20, second largest topic)
- Herfindahl index across agencies: 0.017 (19/20, second least concentrated topic)
- Herfindahl index across states: 0.198 (19/20, second least concentrated topic)

Topwords:

- saber, información, puede, gracias, hacer, caso, servicios, alguna, existencias, debe
- to know, information, can, thank you, to do, case, services, some, existence (in stock), should

FREX words:

- the, puede, hacer, espera, antemano, ayuda, quiero, saber, acudir, quisiera
- (english word), can, to do, wait, in advance (as in, I thank you in advance), help, I want, to know, to resort to, I would like

Top Five Agencies (with proportions of requests in this topic)

SECRETARÍA DE EDUCACIÓN PÚBLICA	0.055
SECRETARÍA DE HACIENDA Y CRÉDITO PÚBLICO	0.038
INSTITUTO MEXICANO DEL SEGURO SOCIAL	0.036
SECRETARÍA DE SALUD	0.033
PROCURADURÍA GENERAL DE LA REPÚBLICA	0.028

Examples

December 17, 2006: INSTITUTO NACIONAL DE LAS MUJERES

- Necesito saber las dependencias que ayudan a las madres solteras en el estado de Nuevo Leon y que ayuda es la que proporcionan. Gracias.
- (Where to go for assistance for single mothers in Nuevo Leon and what help they provide.)

January 17, 2007: SECRETARÍA DE ECONOMÍA

- Les pido de la manera mas atenta me informen adonde acudir y la forma en la que puedo solicitar un credito para iniciar una microempresa con giro en alimentos (panaderia y reposteria) y cuales son los requisitos a cubrir
- (How to apply for a small business loan to open a bakery and what are the requirements for applying.)

February 3, 2009: ADMINISTRACIÓN PORTUARIA INTEGRAL DE ALTAMIRA S.A. DE C.V.

- Un saludo cordial Me permito contactarlos con el motivo de obtener informacion referente a los requisitos que necesitaria cumplir para importar llantas chinas a Mexico me comenta la compañía china que necesitaria checar la NOM no se si ustedes me pudieran ayudar en este sentido sin mas por el momento me despido.
- (Requirements for importing Chinese tires.)

March 20, 2013: PROCURADURÍA GENERAL DE LA REPÚBLICA

- Buenas tardes estoy realizando mis tramites de apadrinamiento y me piden una documento que se llama INFORME DE DATOS REGISTRALES podrian usted proporcionarme la informacion adecuada de los documentos a presentar
- (Documents required for sponsorship process.)

Topic 17: Budgets and Spending

This topic concerns budgets and spending by government agencies. This is reflected in the topwords for this topic, and in the frequency of requests directed to the office of the President and the Ministry of Finance and Public Credit. This is the third largest topic, accounting for 8 percent of all requests. This topic falls below the median values of concentration, both in terms of geography and target agency, reflecting that this is an area of interest to relatively broad areas of the country (beyond just the capital), and across many different agencies.

This topic's share of all requests has been relatively stable, with no clear time trend. A notable spike in requests in this topic appears (see Figure 4 in the main paper) in 2003 shortly after the initial implementation of the access to information law in Mexico. This may reflect a burst of activity requesting previously inaccessible budget and spending information.

Most individual requests in this topic ask for budget allocations or amounts of spending. Some ask for such information in aggregate terms (such as asking for the full budget breakdown for a particular agency), while others are very specific (such as asking how much was spent for a particular use). Most of these requests appear *not* to be made by experts or individuals with high levels of familiarity with government documents or procedures. However, another frequent focus of many of these requests is on spending on advertising or promotional campaigns. Such spending is often the subject of critical attention, particularly as it may be misused for campaign purposes. For example, in the first year of Enrique Peña Nieto's presidency, the federal government exceeded by close to 50 percent the budget that had been approved by congress for official publicity, spending 7.6 billion pesos (about US \$600 million).¹⁸ However, the topic model also captures some unrelated requests, often for "numbers of things" not directly related to budgets or spending, such as tourists or permits.

Likely requesters may include ordinary citizens, civil society groups, journalists, or even politicians or officials themselves—effectively anybody interested in budget or spending information but without (or unaware of) alternative means to access such information. This topic has a high potential for accountability-seeking uses, including any attempt to identify improper or irregular spending, compare or criticize budget priorities, or identify corruption. For some requests, it is also conceivable that government contractors or service providers could have private interests in information on government spending.

Comparative Measures

- Percent of requests: 8.0% (3/20, third largest topic)
- Herfindahl index across agencies: 0.023 (17/20, fourth least concentrated topic)
- Herfindahl index across states: 0.258 (14/20)

Topwords:

- cuanto, año, presupuesto, gastos, monto, cada, destino, solicitado, información, parte
- how much, year, budget, expenses, amount, each, destination, I request, information, part

¹⁸ See: <http://www.fundar.org.mx/mexico/pdf/LibertadDigitalOk.pdf> (p. 61)

FREX words:

- gastos, viajes, viaticos, presupuesto, destino, donativos ,biblioteca, presidente, dinero, spots
- expenses, travel, travel allowance/per diem, budget, destination, donations, library, president, money, ads

Top Five Agencies (with proportions of requests in this topic)

PRESIDENCIA DE LA REPÚBLICA	0.082
SECRETARÍA DE HACIENDA Y CRÉDITO PÚBLICO	0.069
SECRETARÍA DE EDUCACIÓN PÚBLICA	0.06
SECRETARÍA DE SALUD	0.031
SECRETARÍA DE GOBERNACIÓN	0.028

Examples

February 8, 2005: CONSEJO NACIONAL PARA LA CULTURA Y LAS ARTES

- Distribución del presupuesto anual del Conaculta de manera porcentual en todos los rubros (difusión adminstración becas subsidios etc.)que comprende. En los años de 2004 2003 y 2002.
- (Breakdown of annual budget for the National Council for Culture and Arts, 2002-2004)

November 11, 2006: PRESIDENCIA DE LA REPÚBLICA

- Quiero saber cuánto dinero ha gastado la Presidencia de la República en publicidad en el periodo del 1 de enero de 2006 al 31 de octubre de 2006. Esto es en anuncios en los diferentes medios: periódicos revistas radio y televisión.
- (Office of the President spending on print, radio, and television advertising.)

December 13, 2011: SECRETARÍA DE LA DEFENSA NACIONAL

- Gasto de Sedena para bebidas alcohólicas en lo que va del sexenio. La información se solicita por año monto y tipo de bebida adquirida (vinos cerveza tequila etc etc).
- (Defence Ministry spending on alcoholic beverages, broken down by year, amount, and type of beverage)

July 1, 2013: SECRETARÍA DE EDUCACIÓN PÚBLICA

- Solicito por favor los montos otorgados por concepto de subsidio a todas las universidades públicas del país en el presente año fiscal y los criterios utilizados para la asignación del mismo.
- (Subsidy amounts to public universities, and criteria for allocation)

Topic 8: Commercial Information

This topic primarily concerns information of commercial relevance, such as trademarks, business licenses and registrations, product safety approvals, and administrative proceedings or court cases involving specific businesses or products. The topwords for this topic largely reflect words used to refer to specific documents or types of documents. The two agencies that comprise the largest proportions of requests in this topic are COFEPRIS, which manages health and safety regulations for food and pharmaceutical products, and the Mexican Institute of Industrial Property, which manages patents and trademarks. This is the fourth largest topic, accounting for 7.9 percent of all requests. This topic falls below the median in terms of concentration across agencies, but above the median in terms of concentration across states, reflecting a geographical focus in the Federal District.

This topic's share of all requests has generally trended upwards over time, albeit with a temporary dip around 2012. This upward trend may reflect the extent to which commercial requesters are increasingly aware of information requests as a useful tool. Indeed, commercial requesters are among the most frequent users of access-to-information tools in other contexts, such as the United States and the United Kingdom (Worthy 2013, Kwoka 2016).

Most individual requests in this topic request specific documents or statistics regarding commercial documents or processes such as trademarks, product approvals, and business registrations. Many in particular concern safety approvals for chemicals or medical products, or ask if there have been any complaints or administrative proceedings associated with particular brands. However, the topic model also captures some unrelated requests that make reference to similar specific types of documents, but without clear commercial relevance.

The most likely requesters of these types of information are commercial entities themselves, who seek information on their own regulatory processes, or on those of their competitors or potential business partners. This would parallel similar uses in other contexts like the United States, where it is well documented that pharmaceutical companies are frequent users of the Freedom of Information Act, filing requests both to track their own approvals with the Food and Drug Administration and other agencies, as well as to gain information on those of their competitors. It is also possible that some requests in this topic are accountability seeking, particularly in cases of investigations into unsafe products or potential corruption in regulatory processes.

Comparative Measures

- Percent of requests: 7.9% (4/20, fourth largest topic)
- Herfindahl index across agencies: 0.027 (14/20)
- Herfindahl index across states: 0.334 (8/20)

Topwords:

- solicito, copia, documentos, fecha, información, solicitud, numero, oficio, registro, certificada
- request, copy, documents, date, information, application, number, purpose/profession, register, certified

FREX words:

- simple, certificada, oficio, expediente, copia, folio, permiso, expediente, sanitario, sorteos
- single, certified, purpose/profession, file, copy, file, permission, file, health, raffle

Top Five Agencies (with proportions of requests in this topic)

COMISIÓN FEDERAL PARA LA PROTECCIÓN CONTRA RIESGOS SANITARIOS	0.079
INSTITUTO MEXICANO DE LA PROPIEDAD INDUSTRIAL	0.075
SECRETARÍA DE MEDIO AMBIENTE Y RECURSOS NATURALES	0.049
INSTITUTO MEXICANO DEL SEGURO SOCIAL	0.034
SECRETARÍA DE LA FUNCIÓN PÚBLICA	0.034

Examples

January 5, 2005: COMISIÓN FEDERAL DE COMPETENCIA

- Por este medio solicito una copia simple del oficio [*file number*] expediente [*file number*] emitido por la Comisión Federal de Competencia.
- (Requesting copy of a specific file from the Federal Competition Commission.)

March 23, 2008: INSTITUTO MEXICANO DE LA PROPIEDAD INDUSTRIAL

- Informar sobre posibles procedimientos contenciosos administrativos que involucren a la marca y/o aviso comercial [*company name and registration number*] si hubiere algún procedimiento contencioso administrativo el número de expediente del mismo y estado en que se encuentra dicho procedimiento contencioso
- (Asking if there are any administrative proceedings involving a specific company, and if so the file number and state where they are located.)

April 1, 2011: INSTITUTO MEXICANO DE LA PROPIEDAD INDUSTRIAL

- Solicito me proporcione por esta vía el número de solicitudes de registro de marca presentadas antes ese H. Instituto Mexicano de la Propiedad Industrial el día 12 de Noviembre del 2010
- (Number of trademark applications filed on a specific date.)

September 3, 2014: COMISIÓN FEDERAL PARA LA PROTECCIÓN CONTRA RIESGOS
SANITARIOS

- Solicito a la COFEPRIS con respecto al trámite [*case number*]: a) El nombre de la empresa o persona que ingresó el trámite b) El dispositivo médico o producto del cual se pretende obtener un registro sanitario c) La fecha de ingreso del trámite d) La fecha en la que concluyó el trámite (en dado caso de que no haya concluido el trámite favor de especificar que sigue en proceso). e) El sentido de la conclusión del trámite (se otorgó el registro no se otorgó el registro sigue en proceso el trámite). f) La fecha en la que se notificaron prevenciones y la fecha en la que se desahogaron las prevenciones. Requiero que la información de tal forma que se de respuesta a cada uno de los incisos que integran mi solicitud. Solicito la información tal y como la tenga la Cofepris. Si esta información está contenida en varios documentos solicito todos esos documentos.

- (Information associated with a specific product safety registration, including company name, dates, status, and elements missing for the application, noting that if the information is contained across multiple documents, they request all of them.)

Topic 9: Distributive Programs

This topic's focus is on federal programs that distribute funds to individuals, small businesses, and civil society organizations, devoted to issues such as social development, health, agriculture, tourism, and security. The topwords for this topic reflect a focus on projects, programs, the distribution of funds and resources, and development. The agency most frequently receiving requests in this topic is also the Ministry of Social Development, followed by those focusing on health and on agriculture and rural development. This topic is the fifth largest, accounting for 7.1 percent of all requests. It falls below the median in terms of concentration across both states and agencies, reflecting the relatively broad and diffuse nature of requests of this type. This topic's share of all requests has remained relatively stable over time, with several upward swings at various points, but generally reverting to a stable mean proportion afterwards.

Individual requests in this topic generally ask for either process-oriented or outcome-oriented information related to specific distributive programs. Process-oriented requests include asking for lists of outstanding calls for proposals relevant to specific topics, rules of operation and decision-making procedures, requirements for access, or what government agencies participate in a program. Outcome-oriented requests include asking for lists of program beneficiaries or funding awardees, asking for spending figures, or post-hoc program evaluations. However, this topic also captured some unrelated requests, particularly a series of requests sent to multiple government agencies inquiring as to their social media presence – the word “social” in this case likely misled the topic model.

Likely requesters of information about distributive programs include applicants and potential applicants seeking funding or program access, as well as intermediaries who help potential applicants access these types of programs. This topic also has potential for accountability-seeking uses, as journalists, civil society members, or even political actors may seek oversight of distributive programs that may be used for patronage purposes or misuse of government resources. Such requests may seek evidence of irregularities or favoritism in both procedures and outcomes.

Comparative Measures

- Percent of requests: 7.1% (5/20, fifth largest topic)
- Herfindahl index across agencies: 0.024 (16/20)
- Herfindahl index across states: 0.283 (11/20)

Topwords:

- programa, proyecto, desarrollo, nacional, apoyo, social, realizer, recursos, cual, acciones
- program, project, development, national, support/subsidy, social, carry out, resources, which, actions

FREX words:

- estrategia, programa, rural, indígenas, desarrollo, prevenciones, fomento, metros, poblaciones, cooperación
- strategy, program, rural, indigenous, development, prevention, promotion, meters, populations, cooperation

Top Five Agencies (with proportions of requests in this topic)

SECRETARÍA DE DESARROLLO SOCIAL	0.072
SECRETARÍA DE SALUD	0.064
SECRETARÍA DE AGRICULTURA GANADERÍA DESARROLLO RURAL PESCA Y ALIMENTACIÓN	0.058
SECRETARÍA DE EDUCACIÓN PÚBLICA	0.044
SECRETARÍA DE ECONOMÍA	0.043

Examples

May 22, 2008: INSTITUTO NACIONAL DE DESARROLLO SOCIAL

- Padrón o lista de los integrantes de las comisiones dictaminadoras (nombre estado parte de la comisión e institución) de los proyectos presentados a la convocatoria de Asistencia social (AS) del Programa de Coinversión Social del Indesol del año 2007. La existencia de estos documentos se menciona en el primer párrafo del numeral 5.5 de las Reglas de Operación del Programa de Coinversión Social del Indesol publicadas el 28 de febrero de 2007.
- (List of committee members selecting projects for a social investment program in 2007. Reference the paragraph of the Rules of Operation that stipulate the existence of this document.)

January 10, 2010: SECRETARÍA DE DESARROLLO SOCIAL

- convocatorias de proyectos dirigidas a las asociaciones civiles para apoyar a grupos vulnerables pobreza salud educacion y demas referidos al apoyo social en la republica Mexicana
- (Calls for projects by civil associations to support health, education, and vulnerable groups.)

November 4, 2012: SECRETARÍA DE HACIENDA Y CRÉDITO PÚBLICO

- Solicito Información respecto del FORTAMUN (Fondo de Fortalecimiento Municipal): 1.- Cuáles son las Reglas de Operación? 2.- Cómo opera dicho Fondo? 3.- Cuáles son los Proyectos que se financian con dicho Fondo? 4.- Que dependencias participan dentro del Fondo y cuál es su función o nivel de participación? 5.- Cuáles son los requisitos para tener acceso a dicho Fondo?
- (Information about a federal program for municipal development: rules of operation; requirements to access; which ministries participate.)

June 22, 2015: SECRETARÍA DE SALUD

- Nombre de todas las organizaciones que fueron seleccionadas con uno o dos proyectos especificando cuantos por proyectos por organización para recibir financiamiento de la convocatoria pública para la implementación de estrategias de prevención focalizada del VIH y otras ITS 2015
- (Names of all orgs. selected to receive funding from a HIV/STDs prevention program.)

Topic 16: Military, Police, and Crime

This topic's focus is on the military and the criminal justice system. While the top words by highest probability reflect primarily words pertaining the types of information being requested—such as annual statistics and particular cases—the most frequent and exclusive words reflect an emphasis on police, the military, criminals, and drug trafficking. The agencies most frequently receiving requests in this topic are the Attorney General's Office, the Defense Ministry, the Ministry of Public Security, and the Federal Police. This topic is the sixth largest, accounting for 6.8 percent of all requests. It falls above the median in terms of concentration across agencies—reflecting a particular focus on the first two agencies mentioned above—but below the median in terms of concentration across states, reflecting a relatively broad geographic distribution of demand.

This topic's share of all requests remained relatively stable over time (although with several upwards spikes) until roughly 2010, when it began to trend upwards (albeit with a dip again in 2013), perhaps as a result of intensification of drug war-related violence.

Individual requests in this topic most often asked for aggregate statistics on crime, prosecutions, and violence, but several requests also related to security policies. Requests for statistics asked about the number of victims and number of prosecutions for specific crimes. Requests about security policies queried such topics as state treatment of human rights complaints or troop deployment in confrontations with organized crime. For instance, one request asked about the number of human rights victims registered by a human rights NGO had been recognized as victims by the newly formed Executive Commission for Attention to Victims. This commission was created in 2013 at the pressure of social groups representing victims of human rights abuses related to Mexico's drug war and their families.¹⁹ However, this topic also captured some requests in other topics that asked for aggregate statistics, likely included in this topic because they used terms such as “how much” and mentioned dates or geographic units.

Likely requesters of information about the military and security are journalists reporting on crime and the drug war, as well as NGOs that work with victims or local security. This topic has a high potential for accountability as such groups may use the information in campaigns to pressure the state to enhance security, redress victims, or to follow criminal law procedures. Information may be particularly relevant to electoral campaigns, as success and failings in the area of security are of high political salience, particularly in the context of scandals such as the 2014 disappearance of 43 students in the rural municipality of Ayotzinapa, for which protestors have attributed responsibility to state actors, and which is also the focus of several requests in this topic.²⁰

Comparative Measures

- Percent of requests: 6.8% (6/20)

¹⁹ See: <http://expansion.mx/nacional/2012/04/25/la-ley-general-de-victimas-un-logro-mas-para-el-movimiento-por-la-paz>

²⁰ See: <http://www.proceso.com.mx/385551/fue-el-estado-clamor-en-las-calles-del-pais-por-desaparecidos-de-ayotzinapa>

- Herfindahl index across agencies: 0.058 (8/20)
- Herfindahl index across states: 0.275 (13/20)

Topwords:

- cuanto, año, numero, solicito, información, personal, cual, federal, fecha, caso
- how much, year, number, I request, information, personal, which, federal, date, case

FREX words:

- policía, detenidos, decomisadas, militar, narcotráfico, victimas, delitos, migratoria, crimen, Roberto
- police, detained, confiscated, military/soldier, drug trafficking, victims, crimes, migratory, crime, Roberto

Top Five Agencies (with proportions of requests in this topic)

PROCURADURÍA GENERAL DE LA REPÚBLICA	0.179
SECRETARÍA DE LA DEFENSA NACIONAL	0.113
SECRETARÍA DE SEGURIDAD PÚBLICA	0.046
POLICÍA FEDERAL ANTES POLICÍA FEDERAL PREVENTIVA	0.045
INSTITUTO NACIONAL DE MIGRACIÓN	0.044

Examples

March 10, 2006: SECRETARÍA DE LA DEFENSA NACIONAL

- ¿Cuántas armas prohibidas fueron decomisadas en México en el año 2005? ¿Cuántas armas largas y qué tipos fueron decomisadas en México en el año 2005? ¿Desglose de en qué estados fueron decomisadas estas armas?
- (How many banned weapons were seized in Mexico in 2005, broken down by type and state.)

March 15, 2010: PROCURADURÍA GENERAL DE LA REPÚBLICA

- Solicito se me proporcione el número de secuestros que la PROCURADURÍA GENERAL DE LA REPÚBLICA tiene detectado que se han presentado en los últimos cinco años en Hidalgo La información que se solicita se pide que incluya el municipio que se presentó el secuestro así como mes en que se cometió el delito. Asimismo se pide que la información contenga el dato de que si fue mujer o hombre el afectado y la edad en que tenía el momento del secuestro. Muchas gracias
- (Number of kidnappings in the last five years in Hidalgo, broken down by municipality, month, gender, and age.)

September 17, 2014: SECRETARÍA DE GOBERNACIÓN (INCLUYE LA ENTONCES SECRETARÍA DE SEGURIDAD PÚBLICA)

- quiero conocer los nombres con los que identifican los grupos de delincuencia organizada que actualmente operan en Baja California y los subgrupos bandas o pandillas que operan en esta entidad federativa

- (Names of organized crime groups and subgroups currently operating in Baja California.)

March 27, 2015: SECRETARÍA DE LA DEFENSA NACIONAL

- Solicito información sobre los enfrentamientos en que ha participado el Ejército en Iguala Guerrero y sus alrededores de 2010 a la fecha. Se requiere el saldo y circunstancias de esos enfrentamientos.
- (Information on armed clashes in Iguala, Guerrero since 2010.)

Topic 4: Government Employees 1: Salaries and Benefits

This topic concerns labor and employment policies for government employees, particularly information about wages and salaries, pensions, and other benefits. These subjects are clearly reflected in the topwords for this topic, as well as in the two agencies receiving the most requests in this topic: IMSS, the Mexican Social Security Institute, and ISSSTE, the Institute for Social Security and Services for State Workers. This is the seventh largest topic, accounting for 6.7 percent of all requests. This topic falls right in the middle in terms of concentration across agencies, but is the third least concentrated across states, reflecting the relatively broad geographic dispersion of such requests. This topic's share of all requests has been relatively stable over time, with no clear time trend.

Most individual requests in this topic ask for aggregated statistics—pertaining to a specific agency, position, staff level, or other category of government employee—about salaries, wages, benefits, pensions, or other terms of employment. Some requests also ask for information about employment openings. Likely requesters of this topic of information are government employees themselves, seeking information about their own terms of employment or those of their coworkers or superiors. These requests could be particularly useful to individuals involved in disputes over pay or benefits. Other potential requesters are individuals seeking government employment.

The likely uses of requests for this topic are primarily private in nature. While it is possible that this topic could include requests for accountability-seeking purposes, such as seeking oversight of spending on public employees, we consider this unlikely for the type of requests primarily captured by this topic. Accountability-seeking requests regarding public employment would be much more likely to target specific individuals or more discretionary forms of spending, and indeed these appear in other topics we identify.

Comparative Measures

- Percent of requests: 6.7% (7/20)
- Herfindahl index across agencies: 0.045 (10/20)
- Herfindahl index across states: 0.229 (18/20, third least concentrated)

Topwords:

- trabajo, solicito, año, personal, sueldo, social, seguridad, salario, pago, base
- work, request, year, personal, wage, social, security, salary, payment, base (permanent employee)

FREX words:

- jubilación, pensiones, sindicato, tabulador, sueldo, salario, trabajo, percepciones, salarial, nomina
- retirement, pensions, labor union, tabulator, wage, salary, work, payment, salary related, payroll

Top Five Agencies (with proportions of requests in this topic)

INSTITUTO MEXICANO DEL SEGURO SOCIAL

0.147

INSTITUTO DE SEGURIDAD Y SERVICIOS SOCIALES DE LOS TRABAJADORES DEL ESTADO	0.106
SECRETARÍA DE EDUCACIÓN PÚBLICA	0.053
SECRETARÍA DE HACIENDA Y CRÉDITO PÚBLICO	0.05
SECRETARÍA DEL TRABAJO Y PREVISIÓN SOCIAL	0.033

Examples

January 23, 2004: SECRETARÍA DE LA DEFENSA NACIONAL

- En la SEDENA en 2003: ¿Cuál fue el % de incremento al salario base del personal de mando? ¿Cuál fue el % de incremento a la compensación garantizada? ¿Cuál fue el % de incremento a prestaciones? ¿Cuántos empleados se beneficiaron por estos incrementos?
- (In 2003 for the Ministry of Defense command staff, what were the increases to base salary, guaranteed compensation, and benefits, and how many employees benefitted from these?)

October 14, 2008: PEMEX EXPLORACIÓN Y PRODUCCIÓN

- SALARIOS BASE Y PRESTACIONES DESGLOSADAS INCLUYENDO BONO AL DESEMPEÑO DEL PERSONAL DE CONFIANZA DEL NIVEL 34 35 37 39 y 41 CORRESPONDIENTES A LOS AÑOS 200 2001 2002 2003 2004 2005 2006 2007 Y 2008
- (Base salary, benefits, and bonuses for staff levels 34, 35, 37, 39 and 41 for years 2000 through 2008.)

April 9, 2012: INSTITUTO DE SEGURIDAD Y SERVICIOS SOCIALES DE LOS
TRABAJADORES DEL ESTADO

- POR MEDIO DEL PRESENTE ME PERMITO SOLICITAR TABULADOR DE SUELDOS SELLADO Y CERTIFICADO DE LOS AÑOS: 2009 2010 NY 2011 PUESTO: M02034 DESCRIPCION DEL PUESTO: ENFERMERA ESPECIALISTA A NIVEL Y SUBNIVEL: 08/6 REGION: 2/0
- (Salary schedule for years 2009 to 2011 for a specific level of nurse specialist in a specific region.)

November 25, 2013: INSTITUTO MEXICANO DEL SEGURO SOCIAL

- Solicito al Instituto las estadísticas del salario promedio de cotización de los trabajadores afiliados (LSS-97) al seguro de Retiro Cesantía en Edad Avanzada y Vejez por edad así como el número de semanas promedio de cotización.
- (Average base wage of unionized workers, retirement, unemployment insurance, and average weekly contributions.)

Topic 5: Government Employees 2: Functions and Qualifications

This topic primarily concerns information about government employees, but with a distinct focus from the preceding topic. Requests in this topic tend to concern job descriptions of particular positions, organization charts and contact lists of agencies, and the qualifications of particular employees, often with high-ranking positions. While the topwords for this topic largely reflect words pertaining to organizations and positions, the example requests make clearer the specific types of information involved. Additionally, the agency making up the highest proportion of requests in this topic is the Secretaría de la Función Pública (Ministry of Public Administration, similar to an office of Audit and Control), several concerning complaints or investigations regarding government employees. This is the eighth largest topic, making up 6.3 percent of all requests. This topic falls below the median in terms of agencies, and notably is more dispersed across agencies compared to Topic 4, which was more concentrated in agencies handling social security. However, this topic is the fifth most concentrated topic across states, reflecting a relatively high concentration of requests coming from the Federal District.

This topic's share of all requests has trended downwards over time, albeit with a shift upwards in late 2012 and early 2013, followed by another downward trend. The downward trend over time is consistent with a decreasing need of individuals to obtain these types of information via formal request, as more government agencies make information about organization charts, contact information, and individual employees proactively available on the internet. The spike in 2012-2013 is likely a response to changes in agency management following the national change in power with the election of Enrique Peña Nieto, who took office on December 1, 2012.

Most individual requests in this topic seek information about the identities, contact information, functions, or qualifications of specific positions in government agencies. Examples include asking for the names and qualifications of the top-level positions of an agency, asking for the organization chart or official contact lists of an agency, asking for the name of the official who handles a specific task, or asking for the procedures and requirements for employment in a specific position. Some requests specifically ask for the resumes of individual officials. Finally, some requests ask for information about complaints or investigations regarding government employees. Distinct from Topic 1 (which will be described subsequently), however, very few of any of these requests focus on specific individuals by name; rather they focus on particular official positions and roles.

Likely requesters of these types of information include individuals trying to identify the correct government official relevant to their particular concern or need, trying to identify how to contact a particular official, or seeking government employment themselves. Many requests in this topic also have potential to be accountability-seeking, by attempting to identify unqualified government officials or patronage hiring practices, or seeking information about complaints or investigations regarding government employees.

Comparative Measures

- Percent of requests: 6.3% (8/20)
- Herfindahl index across agencies: 0.027 (13/20)
- Herfindahl index across states: 0.358 (5/20, fifth most concentrated topic)

Topwords:

- pública, administración, nombre, direcciones, general, area, servidores, organism, funcionarios, cargo
- public, administration, name, addresses, general, area, (public) servants, organization, functionaries, job

FREX words:

- Villagomez, organism, servidores, internacional, funcionarios, vacantes, funcionarios, puesto, carrera, cargo
- Villagomez, organization, employees, international, functionaries, vacancies, functionaries, position, career/degree, job

Top Five Agencies (with proportions of requests in this topic)

SECRETARÍA DE LA FUNCIÓN PÚBLICA	0.124
SECRETARÍA DE EDUCACIÓN PÚBLICA	0.048
SERVICIO DE ADMINISTRACIÓN TRIBUTARIA	0.033
SECRETARÍA DE HACIENDA Y CRÉDITO PÚBLICO	0.033
SECRETARÍA DE SALUD	0.032

Examples

November 14, 2007: AEROPUERTO INTERNACIONAL DE LA CIUDAD DE MÉXICO S.A. DE C.V.

- Solicito nombre y perfil académico de: Coordinador de Archivos Responsables de archivo de trámite Responsable de archivo de concentración y archivo histórico Personal operativo que tiene funciones de archivista exceptuando los anteriormente mencionados.
- (Names and academic qualifications for the archives coordinator and individuals with various other archival roles.)

May 17, 2010: SECRETARÍA DE LA FUNCIÓN PÚBLICA

- ¿Qué órganos centralizados descentralizados o del sector paraestatal se encuentran sujetos al Servicio Profesional de Carrera en la Administración Pública Federal?
- (What centralized, decentralized, or parastatal entities are subject to the Professional Career Service?)

February 22, 2012: SECRETARÍA DE EDUCACIÓN PÚBLICA

- solicito versión pública de los títulos profesionales y todos y cada uno de los subdirectores de área directores de área directores generales directores generales adjuntos subsecretarios y jefes de departamento de la dependencia.
- (Public versions of qualifications for all of the agency's directors, subdirectors, director generals, undersecretaries, and heads of department.)

January 29, 2014: CENTRO NACIONAL DE METROLOGÍA

- La dirección de correo electrónico respecto de todos los funcionarios empleados prestadores de servicios y toda clase de servidores públicos que actualmente se desempeña tanto en esa dependencia como en las unidades administrativas que dependen de la misma. Cabe aclarar que se está solicitando el correo electrónico que la dependencia gubernamental asignó a los empleados prestadores de servicios y toda clase de servidores públicos que actualmente se desempeñen en esa dependencia; en ningún caso se está solicitando información personal de algún empleado en particular.
- (Email addresses of all of the agency's officials, employees, and public servants. Emphasizes that the request is for the email addresses assigned to employees, and not for the personal information of any particular employee.)

Topic 12: Energy and Public Utilities

Requests in this topic concern energy and public utilities – oil and gas production and consumption, electricity production, and provision of public utilities like water, electricity, and lighting. These are reflected both in the top words, and in the most common agencies receiving this topic of request—the Federal Electricity Commission, the Secretary of Communication and Transportation, and entities of PEMEX, the state oil company. This is the ninth largest topic, accounting for 5.6 percent of all requests. It falls below the median in concentration across agencies and is the fourth least concentrated across states, reflecting geographically dispersed interest. This topic’s share of requests has remained mostly stable over time, though with a very slight downward trend.

Most individual requests in this topic ask for statistics and figures regarding the production and consumption of oil, gas, electricity and water, as well as regarding the provision of public utilities like electricity, water, lighting, and transportation infrastructure. Some of these focus on specific areas, such as requesting statistics about the provision of public street lighting in specific municipalities. Some requests also pertain to the costs of extracting oil or production figures for particular locations, lists of oil projects, or specific questions about the process of production and distribution of oil and gas. Others seek details on incidents of theft of oil or natural gas.

Likely requesters of this topic of information may be varied. They include journalists, investors, government officials themselves, and ordinary citizens. Purposes could be private, such as pertaining to the oil industry or carrying out private research. However, many requests in this topic are likely accountability-seeking with goals either of attempting to monitor or identify corruption in the energy sector, or of evaluating government performance in providing public services, or the results of government spending on infrastructure projects. PEMEX, the state oil company is often criticized for wasteful spending, corrupt hiring practices, and concessions practices, which became a highly salient political issue in the context of a 2013 reform enabling private contracting for the national oil company.²¹

Comparative Measures

- Percent of requests: 5.6% (9/20, largest topic)
- Herfindahl index across agencies: 0.04 (12/20)
- Herfindahl index across states: 0.254 (17/20, fourth least concentrated)

Topwords:

- pemex, mexicano, servicios, año, empresa, operación, tipo, electrónico, transporte, producciones
- PEMEX, Mexican, services, year, business, operation, type, electronic, transportation, production

FREX words:

²¹ See: <http://cidac.org/petroleos-mexicanos-a-la-deriva/>

- alumbrado, gastos, tarifa, energía, diesel, pemex, cfe, telecomunicaciones, refinación, exportación
- lighting, expense, price/rate, energy, diesel, PEMEX, CFE (electrical utility), telecommunications, refining, exportation

Top Five Agencies (with proportions of requests in this topic)

COMISIÓN FEDERAL DE ELECTRICIDAD	0.119
PEMEX EXPLORACIÓN Y PRODUCCIÓN	0.083
SECRETARÍA DE COMUNICACIONES Y TRANSPORTES	0.073
PETRÓLEOS MEXICANOS	0.055
PEMEX REFINACIÓN	0.054

Examples

June 20, 2005: COMISIÓN NACIONAL DEL AGUA

- Producción de Agua volúmenes y procedencia (pozos) que proveen al municipio de Guaymas Sonora Consumo (usos) de Agua (industrial comercial agrícola y doméstica) tanto nacional como para el municipio de Guaymas Sonora
- (Water production volume and sources for Guaymas, Sonora, as well as water consumption and type of use, both nationally and for Guaymas, Sonora.)

October 23, 2008: PEMEX EXPLORACIÓN Y PRODUCCIÓN

- Costo de extracción unitario o promedio o estimado de un barril de crudo maya olmeca istmo y mezcla mexicana por zona territorial o regional o por pozo. Los costos de extracción deben ser mensuales y/o anuales de 2008 hasta 1979 (o hasta donde se tengan registros.
- (Extraction costs of oil, broken down by region and month/year from 1979 to 2008.)

February 18, 2013: PEMEX EXPLORACIÓN Y PRODUCCIÓN

- Con motivo de la elaboración del Balance Energético del estado de Hidalgo que realiza la Benemérita Universidad Autónoma de Puebla (BUAP) me dirijo a ustedes para solicitar la estadística anual correspondiente al periodo 2004- 2011 para el estado de Hidalgo referente a: La cantidad de petróleo crudo (en barriles) condensados (en barriles) y gas natural (en metros cúbicos) que sale del estado de Hidalgo para su transformación indicando las plantas de destino y su ubicación.
- (Annual statistics for Hidalgo for 2004 to 2011 on the amounts of crude oil, condensates, and natural gas leaving Hidalgo for processing plants, including destination and location.)

August 22, 2012: COMISIÓN FEDERAL DE ELECTRICIDAD

- 1.-Censo de Luminarias (Lámparas de Alumbrado Público) que incluya las potencias y la tecnología en el municipio de Morelia Michoacán. 2.-Copia del Convenio de Recaudación y Administración del Derecho de Alumbrado Público que tenga suscrito con el Municipio de Morelia Michoacán. 3.-Cantidad de servicios medidos y no medidos en el Alumbrado Público del Municipio de Morelia Michoacán. 4.-Historial de Consumos y

facturación de Energía Eléctrica del Alumbrado Público de los últimos 3 (tres) años del Municipio de Morelia Michoacán. 5.-Facturación por consumo de energía eléctrica del Alumbrado Público de los últimos 3(tres) años del Municipio de Morelia Michoacán. 6.- Facturación por consumo de energía eléctrica de los últimos 3 (tres) años de El Organismo Operador de Agua Potable Alcantarillado y Saneamiento de Morelia Michoacán. 7.-Facturación del consumo de energía eléctrica de los últimos 3 (tres) años de los inmuebles propiedad del Municipio de Morelia Michoacán. 8.-Estados de cuenta de los últimos 3 (tres) años de la facturación del consumo de energía eléctrica del Alumbrado Público del Municipio de Morelia Michoacán. 9.-Cantidades recaudados en los últimos 3(tres) años del Derecho de Alumbrado Público en el Municipio de Morelia Michoacán. (desglosadas por mes). 10.-Cantidad de los remanentes derivados del cobro del Derecho de Alumbrado Público de los últimos 3 (tres) años del Municipio de Morelia Michoacán. 11.- Que funcionarios de la Comisión Federal de Electricidad tienen las facultades para suscribir convenios addendums o acuerdos en relación a la recaudación del Derecho de Alumbrado Público con el Municipio de Morelia Michoacán. 12.- Consumo en Kilowatts por concepto de energía eléctrica de alumbrado público de los últimos tres años del Municipio de Morelia Michoacán. 13.-Consumo en Kilowatts por concepto de energía eléctrica de general de los últimos tres años del Municipio de Morelia Michoacán.

- (Several specific questions regarding public lighting, electricity consumption, and billing, in Morelia, Michoacán.)

Topic 14: Internal Procedures and Official Documents

This is perhaps the most difficult topic to label; the common thread in the requests is references to internal procedures for government agencies and regulations, often referencing specific statutes. The top words include many words common to legislation and official government documents (Article, Disposition, Regulation), as well as references to the ATI law itself (Law, Federal, Access, Federal, Transparent, Governmental). This was the tenth largest topic, accounting for 5.2 percent of all requests. Requests are not specific to any policy area, as reflected in the high dispersion across agencies (third least concentrated) and the most frequently requested agencies for this topic are among the leaders for overall requests (Treasury, Ministry of Public Administration, Ministry of Public Education, Mexican Social Security Institute). However requests were slightly above the median in geographic concentration, with over half of requests coming from the Federal District. This topic's share of requests has remained quite stable over time.

Most requests asked for specific documents, either pieces of legislation or internal agency documents. For instance, one request asked for a code of ethics that the Ministry of Public Administration is required to produce, referencing the specific article of the Federal Law of the Administrative Responsibilities of Public Servants stating this requirement. The fact that requests in this category remain constant as a proportion of all requests suggests that these documents are not accessible online (as was the case for Topic 6). Some requests also asked for justifications of internal agency decisions or legal reforms, such as the reasoning behind the reform of the federal consumer protection law. This topic also captured several requests that would appear to be closely related to other topics owing to the policy areas referenced, but likely were included here because they included several words referencing the federal access to information law.

Likely requesters of this topic may be varied, however are likely to mostly be “insiders” or informed figures (journalists, bureaucrats, firms). Typical citizens would be less likely to know the names of official documents or reference legal language. Some proportion of requests are likely to be accountability seeking, as they allow the requester to gauge the degree to which government agencies comply with official requirements. However the bulk of requests pertain to official statutes rather than data or documents that can be used to gauge government performance.

Comparative Measures

- Percent of requests: 5.2% (10/20)
- Herfindahl index across agencies: 0.022 (18/20, third-least concentrated)
- Herfindahl index across states: 0.309 (9/20)

Topwords:

- ley, pública, federal, información, artículo, acceso, transparente, unidad, gubernamental, conforme
- law, public, federal, information, article, access, transparent, unit/agency, governmental, in accordance

FREX words:

- reglamento, dispuesto, ley, artículo, fracciones, acceso, artículo, disposiciones, constitución, conforme, gubernamental
- regulation, willing/ready, law, article, sections, access, article, disposition/regulation, constitution, in accordance, governmental

Top Five Agencies (with proportions of requests in this topic)

SECRETARÍA DE HACIENDA Y CRÉDITO PÚBLICO	0.057
SECRETARÍA DE LA FUNCIÓN PÚBLICA	0.052
SECRETARÍA DE EDUCACIÓN PÚBLICA	0.043
INSTITUTO MEXICANO DEL SEGURO SOCIAL	0.039
PROCURADURÍA GENERAL DE LA REPÚBLICA	0.036

Examples

August 1, 2006: PROCURADURÍA FEDERAL DEL CONSUMIDOR

- solicito de la manera más atenta me proporcionen la exposición de motivos de la reforma del 4 de febrero de 2004 a la Ley Federal de Protección al consumidor
- (Preamble of the reform of the Federal Consumer Protection Law.)

March 23, 2010: COMISIÓN FEDERAL PARA LA PROTECCIÓN CONTRA RIESGOS SANITARIOS

- SOLICITO EL PROCEDIMIENTO PARA REALIZAR VISITAS DE VERIFICACIÓN Y QUE ES SEÑALADO EN EL ANTEPENÚLTIMO PÁRRAFO DEL ARTÍCULO 167 DEL REGLAMENTO DE INSUMOS PARA LA SALUD.
- (Wants to know the procedure for inspections laid out in a specific article/paragraph of the Regulations on Health Supplies.)

April 25, 2011: COMISIÓN NACIONAL BANCARIA Y DE VALORES

- Con fundamento en el artículo 8 constitucional; fracción XVII del artículo 7 y artículo 9 de la ley federal de transparencia y acceso a la información pública; 1 2 22 23 24 y 25 de la ley del servicio profesional de carrera en la administración pública federal solicito: Tenga a bien informarme en que tiempo se publican las convocatorias para ocupar las plazas del primer nivel en la referida institución; Cules son los criterios que considera el comité para llevar a cabo el procedimiento de selección para ocupar los cargos. Sin más y agradeciendo de antemano la atención que se sirva dar a la presente reciban mis respetos y consideraciones.
- (Based on Article 8 of the Constitution and relevant sections of the transparency law and law on public servants, when are calls posted to fill positions at the first level of the institution, and what are the criteria considered in the selection procedures.)

August 27, 2014: SECRETARÍA DE ECONOMÍA

- Número de solicitudes presentadas desde agosto de 2013 a agosto de 2014 para inscribir actos jurídicos en el Registro Público de Minería que fueron negadas por no cumplir el

requisito previsto en el artículo 78 del Reglamento de la Ley Minera referente a que se presente el acto ratificado sin que pasen 5 días de dicha ratificación sin que haya mediado requerimiento para su subsanación como lo establece el artículo 17-A de la Ley Federal de Procedimiento Administrativo.

- (Number of Public Mining Registry applications submitted between August 2013 and August 2014 that were denied for not meeting specific requirements.)

Topic 13: Medical Statistics

Requests in this topic concerned medical statistics, including incidence of medical conditions and treatment delivered by public clinics. The top words reflect the medical nature of the information sought (patients, mortality, medication) and statistical groupings (number, year, national). The most common agencies for this topic of request are centrally involved in the provision of health services (Social Security Institute, Health Ministry, Social Security Institute for State Employees) or are agencies that offer health care to their own employees (PEMEX, Ministry of Defense). This is the eleventh largest topic, accounting for 4.7 percent of all requests and is highly concentrated across agencies (fifth) and geographically (sixth). This topic's share of requests has slightly increased over time, with notable spikes in 2012, 2013, and 2015, reflecting increased attention to health indicators during the current presidential term.

Individual requests in this topic mostly ask for numerical indicators. Several requests seek information about the incidence of medical conditions, such as the number of children with cystic fibrosis or the number of patients diagnosed with Hepatitis C treated in a given hospital. Some requests pertain to statistics reflecting the actual treatment of conditions, such as requesting the number of people receiving flu vaccines or mortality rates for those with liver failure.

Likely requesters of this topic are journalists or researchers interested in tracking the incidence of medical conditions. Few requests would appear to be useful for private purposes, such as deciding how to pursue treatment for a given medical condition. Most requests would also appear to not be particularly accountability generating, as they do not pertain to spending, personnel, or procedures. However, some may be verifying that agencies accurately report health statistics. In contrast, topics 11 and 20 are much more oriented to tracking purchasing of medical supplies, an issue closely linked to corruption.

Comparative Measures

- Percent of requests: 4.7% (11/20)
- Herfindahl index across agencies: 0.126 (5/20, fifth-most concentrated)
- Herfindahl index across states: 0.346 (6/20, sixth-most concentrated)

Topwords:

- medicamento, año, numero, hospital, solicitado, salud, especial, edad, información, nacional
- medication, year, number, hospital, I request, health, special, age, information, national

FREX words:

- pacientes, hospitalaria, recetas, morbilidad, medicina, hospital, tratamiento, sexo, enfermedades, medicamento
- patients, hospital-related, prescriptions, mortality, medicine, hospital, treatment, sex, illnesses, medication

Top Five Agencies (with proportions of requests in this topic)

INSTITUTO MEXICANO DEL SEGURO SOCIAL	0.304
SECRETARÍA DE SALUD	0.146

INSTITUTO DE SEGURIDAD Y SERVICIOS SOCIALES DE LOS TRABAJADORES DEL ESTADO	0.095
PETRÓLEOS MEXICANOS	0.022
SECRETARÍA DE LA DEFENSA NACIONAL	0.02

Examples

December 6, 2006: INSTITUTO DE SEGURIDAD Y SERVICIOS SOCIALES DE LOS TRABAJADORES DEL ESTADO

- Requiero informe cuantitativo y cualitativo referente a la morbilidad y mortalidad de personas en general por insuficiencia hepática por causas a partir del 2000 hasta la presente fecha en el Estado de Quintana Roo desglosado por municipio grupo etario(rango de edades) y sexo.
- (Information on morbidity and mortality of patients with liver failure in Quintana Roo, broken down by municipality, age, and sex.)

May 20, 2009: INSTITUTO DE SEGURIDAD Y SERVICIOS SOCIALES DE LOS TRABAJADORES DEL ESTADO

- Cuántas personas mayores de veinte años y menores de treinta años acudieron a vacunarse por influenza en la ciudad de Tijuana Baja California en el año de dos mil ocho?
- (How many people ages 20-30 were vaccinated for influenza in Tijuana, Baja California in 2008?)

October 21, 2011: INSTITUTO MEXICANO DEL SEGURO SOCIAL

- Solicito a la Unidad Médica de Alta Especialidad (UMAE) Hospital Pediatría Centro Médico Nacional Occidente Guadalajara en Jalisco la cantidad de niños y niñas diagnosticadas con fibrosis quística y que fueron atendidos en esa UMAE del 01/01/2010 al 31/12/2010.
- (Number of children diagnosed with cystic fibrosis and treated in a specific hospital in 2010.)

January 3, 2013: INSTITUTO MEXICANO DEL SEGURO SOCIAL

- Solicito a la delegación Veracruz Sur del IMSS el número total de personas diagnosticadas con el virus de la hepatitis C del 01/01/2012 al 31/12/2012 desagregadas por unidad hospitalaria edad y sexo. No requiero nombres sólo números consecutivos
- (Asking the southern Veracruz IMSS office how many patients were diagnosed with hepatitis C in 2012, broken down by age and sex.)

Topic 1: Government Employees 3: Specific Personnel

This topic's focus is on information about specific persons, many of whom are government employees or information regarding their official government dealings, such as licensing and contracts. Many requests ask for specific documents, often referenced by number, but the most notable commonality is the mention of names (e.g. José, Luis, González, Pérez), as these constitute most of the top words. The most common agencies that receive these requests are those that have a large number of employees (Education), provide benefits (Social Security), or engage in oversight of personnel practices (Ministry of Public Administration). However concentration across agencies and across states are both below the median. This is the twelfth largest topic, accounting for 4.1 percent of all requests. This topic's share of total requests has remained relatively constant over time, with a slight uptick in 2011.

Individual requests quite often request documents likely to be found in a government employee's internal HR records, such as declaration of assets and CV, or information about specific employees' behavior, such as expense reports or work schedules.

While topic 5 includes many requests that ask about aggregate data on government hiring and benefits, likely useful for those seeking federal government employment, the requests in this topic are more clearly associated with accountability seeking. The information is likely to be requested by journalists, attorneys, or watchdogs investigating patronage hiring, the qualifications of government personnel or corruption.

Comparative Measures

- Percent of requests: 4.1% (12/20)
- Herfindahl index across agencies: 0.025 (15/20)
- Herfindahl index across states: 0.279 (12/20)

Topwords:

- mexicano, licitación, José, general, director, colonia, delegación, copia, colonia, Luis
- Mexican, request for bids, José, general, director, neighborhood, borough, copy, Luis, Gonzalez

FREX words:

- Perez, Huasteca, Ortiz, Vargas, Javier, Castro, primer, Hector, Muñoz, Aguilar
- Perez, Huasteca, Ortiz, Vargas, Javier, Castro, first, Hector, Muñoz, Aguilar

Top Five Agencies (with proportions of requests in this topic)

SECRETARÍA DE EDUCACIÓN PÚBLICA	0.089
INSTITUTO MEXICANO DEL SEGURO SOCIAL	0.057
SECRETARÍA DE LA FUNCIÓN PÚBLICA	0.041
SECRETARÍA DE SALUD	0.033
INSTITUTO DE SEGURIDAD Y SERVICIOS SOCIALES DE LOS TRABAJADORES DEL ESTADO	0.031

Examples:

January 24, 2006: SECRETARÍA DE LA FUNCIÓN PÚBLICA

- SOLICITO COPIA DE LA ÚLTIMA DECLARACIÓN PATRIMONIAL QUE PRESENTÓ [individual name] REPRESENTANTE ESTATAL DE LA SECRETARÍA DE GOBERNACIÓN EN SINALOA ANTE LA SECRETARÍA DE LA FUNCIÓN PÚBLICA.
- (Requesting copy of most recent asset declaration for the Ministry of the Interior's state delegate in Sinaloa.)

February 8, 2007: INSTITUTO PARA LA PROTECCIÓN AL AHORRO BANCARIO

- Copia de las cartas de renuncia de los ex vocales del IPAB [five individual names]
- (Copies of the resignation letters of five former members of the Bank Savings Protection Institute.)

January 8, 2009: SECRETARÍA DE EDUCACIÓN PÚBLICA

- Solicito me informe número de título profesional institución que lo expidió y fecha de expedición así como número y fecha de expedición de la cédula profesional de las siguientes personas: [Long list of individual names]
- (Requesting the professional licenses of a long list of individuals, along with the issuing institution and date of issue.)

June 24, 2011: SECRETARÍA DE EDUCACIÓN PÚBLICA

- deseo informacion sobre los horarios puestos de trabajo de las sig personas: [Long list of individual names] TODOS ELLOS TRABAJADORES DEL INSTITUTO TECNOLÓGICO DE NUEVO LEON (DEGEST)
- (Requesting the work schedules of several employees at the Instituto Tecnológico de Nuevo León.)

Topic 18: Public Procurement 1: Service Providers

Requests in this topic concern government contracting of services or facilities. Most of the requests are concerned with specific contracting or bidding procedures, such as the process by which a government agency chose a contractor for internet service, concessions for public works, or hiring external consultants. Compared to topic 19, which focuses on amounts of money and purchasing of supplies, this topic is largely concerned with selection procedures and the particular actors contracted by government agencies. This was the least concentrated topic across agencies, suggesting that the type of information requested was not particular to any given policy area, although the top requested agencies may be those that invest the most in facilities (PEMEX, Social Security, Ministry of Transportation and Communication, Federal Electricity Commission). However this was the third-most concentrated topic across states, with 65 percent of requests coming from the Federal District, suggesting that many of the interested parties are perhaps service providers located in the capital. This topic constitutes 4.1 percent of all requests, 13th out of the twenty topics. The overall trend of this topic as a portion of all requests is relatively constant, with frequent spikes, perhaps corresponding to the budget cycle.

Individual requests in this topic often relate to hiring of external consultants or contractors to provide services for government agencies, such as a request about the bidding process and budget for contracting celular service for employees in the education ministry and several requests about the contract for an external adviser for insurance submitted to many ministries. Other requests had to do with facilities and public works, such as a request asking PEMEX about the number of contracts that had been given to Brazilian firms for construction and other services. The vast majority of requests reviewed fit these descriptions, but a few exceptional requests may have been included because they referenced agency facilities and/or asked about budgets for maintenance and other operations.

This topic is of mixed utility for accountability seeking. Many requesters in this topic are probably providers of the very services that they are asking about. By asking who receives government contracts and the procedures by which they are selected, these providers may be investigating strategies for winning contracts themselves. Other requesters may be journalistic or watchdog organizations, conducting oversight of contracting practices, suggesting that this topic could be quite conducive to social accountability. Scandals regarding the choices of government contractors are quite salient in Mexico, especially in light of the 2014 “Casa Blanca” scandal, in which the federal government was accused of rigging a bidding process for a high-speed rail to favor a personal friend of the president who had granted the presidents’ family use of a 7 million dollar residence.²²

Comparative Measures

- Percent of requests: 4.1% (13/20)
- Herfindahl index across agencies: 0.011 (20/20, the single least concentrated)
- Herfindahl index across states: 0.442 (3/20, third-most concentrated)

²² See: <http://aristeguinoticias.com/0911/mexico/la-casa-blanca-de-enrique-pena-nieto/>

Topwords:

- contrato, servicios, empresa, monto, solicito, nombre, dependents, licitación, fecha, seguridad
- contract, services, firm, amount, i request, name, dependents/employees, request for bids, date, security

FREX words:

- PEF, prestación, maniobras, fletes, contrato, celebrado, poliza, paquetería, empresa, formalizados
- federal budget, benefit, maneuver, freight, contract, celebrated/taken place, policy (insurance), shipping, firm, formalized

Top Five Agencies (with proportions of requests in this topic)

PEMEX EXPLORACIÓN Y PRODUCCIÓN	0.036
INSTITUTO MEXICANO DEL SEGURO SOCIAL	0.034
SECRETARÍA DE COMUNICACIONES Y TRANSPORTES	0.03
COMISIÓN FEDERAL DE ELECTRICIDAD	0.03
PETRÓLEOS MEXICANOS	0.025

Examples

July 8, 2004: SECRETARÍA DE GOBERNACIÓN

- Proporcionar el nombre del asesor externo en materia de seguros vigencia y monto del contrato Indicar el esquema de adjudicacion de dicho contrato (licitacion publica licitacion restringida a cuandos menos tres personas o adjudicacion directa). Proporcionar el nombre y puesto del encargado interno de seguros.
- (Requesting the name of the external insurance consultant and internal insurance manager, amount and type of the contract.)

November 6, 2006: INSTITUTO NACIONAL DE ASTROFÍSICA ÓPTICA Y ELECTRÓNICA

- proporcionar la justificacion del area usuaria de la adjudicacion directa de los servicios de asesor externo en materia de seguros para el ejercico 2005 al 2006
- (Asking for justification of the use of a direct award for selecting an external insurance consultant in 2005-2006.)

March 2, 2011: PEMEX GAS Y PETROQUÍMICA BÁSICA

- Quisiera saber: 1. Cuál es el número de inmuebles por entidad federativa de dicha dependencia. 2. Cuál es la ubicación y domicilio de cada uno de los inmuebles. 3. Cuál es el número de empleados de dicha dependencia en cada uno de los inmuebles. 4. A cuanto ascendió el gasto correspondiente al ejercicio fiscal 2010 para vigilancia seguridad y limpieza de los inmuebles e instalaciones de la misma. 5. A cuánto asciende el presupuesto autorizado por la secretaría de hacienda crédito público a esa dependencia para vigilancia seguridad y limpieza de los inmuebles e instalaciones de la misma para 2011. 6. A cuánto asciende el monto destinado para cada uno de los inmuebles e

instalaciones por concepto de servicios de vigilancia seguridad y limpieza que pertenecen a esta dependencia. 7. Cuáles son los requisitos que solicita la dependencia a las empresas de seguridad vigilancia y limpieza para ser contratadas. 8. Cuál es la modalidad para la contratación de las empresas de seguridad vigilancia y limpieza: adjudicación invitación a cuando menos tres proveedores o licitación. 9. Cuáles son los nombres y domicilios de las empresas que prestan los servicios de vigilancia seguridad y limpieza en cada uno de los inmuebles e instalaciones. 10. Cuantos elementos de vigilancia seguridad y limpieza se tienen contratados para cada una de los inmuebles y/o instalaciones de la institución.

- (Asking several questions regarding the security and cleaning companies hired by the agency for their properties and facilities, including the names and addresses of companies, requirements for hiring, budget allocations and expenditures for security and cleaning services, as well as basic information about the properties.)

March 3, 2013: INSTITUTO NACIONAL DE ANTROPOLOGÍA E HISTORIA

- Inventario de equipo de cómputo (de escritorio o portátil) con el número de licencia de sistema operativo así como la vigencia de esta última la forma como fué adquirido y la versión pública del contrato. Si el proveedor vendió los equipos con licencia o los incluyó en su contrato proporcionar el OEM y el soporte del mismo. Esto se solicita para los ejercicios 2011 2012 y lo que va de 2013. No importa la forma en que se hayan adquirido ya sea por adjudicación directa invitación a cuando menos 3 personas o licitación pública. Si fue una contratación por servicios administrados proporcionar el soporte de las licencias que se utilizan y la vigencia de las mismas.
- (Requesting an inventory of all computer equipment, operating system licenses and how they were acquired, and public versions of the contracts, for years 2011 to 2013. Request applies to direct awards, limited competitions, and fully public tenders.)

Topic 2: Banking, Finance, and Taxes

This topic pertains to financial information, particularly information about tax collection and regulation of financial institutions. The financial nature of the topic is reflected in the top words (e.g., credit, fiscal, bank, treasury) as well as the most requested agencies in this topic: the tax agency, the treasury, and the national banking commission. The measure of concentration across agencies is only slightly above the median, as the bulk of requests are split among several ministries that deal in financial issues. This topic is well below the median in geographic concentration, with slightly less than half of requests originating in the Federal District. This topic constitutes 3.8 percent of all requests, fourteenth out of the twenty topics. The overall trend of requests in this topic as a proportion of total requests is relatively stable, with notable spikes in 2006 and 2015, likely owing to public banking scandals.

Individual requests in this topic mostly are either about taxation or about regulation of the financial sector. These requests tend not to ask for aggregate data, but rather identify particular cases or people. For instance one request identified several specific advertising firms, asking how much money in taxes had been collected from each. Many requests pertaining to the financial sector reference specific investigations or legal processes. For instance, a glut of requests in 2015 sought information related to the FICREA scandal, in which a financial institution was found to have laundered money from savings accounts, leading to a highly publicized investigation by the National Banking Commission.²³ Several requests concerned the resolution of this dispute, particularly whether victims of the fraud would be reimbursed.

This topic has a high potential for accountability, as requests are likely to be oriented to uncovering instances of tax evasion or following up on investigations into financial crimes such as in the FICREA scandal. For such highly publicized cases, requesters were likely journalists or representatives of citizens watchdog groups. Although some requests likely also were made by the ordinary clients of financial institutions. The fourth most highly requested agency for this topic was the Financial Consumer Protection Agency suggesting that a significant portion of requesters were consumers of financial institutions, seeking redress through this agent of horizontal accountability.

Comparative Measures

- Percent of requests: 3.8% (14/20)
- Herfindahl index across agencies: 0.046 (9/20)
- Herfindahl index across states: 0.256 (15/20)

Topwords:

- credito, fiscal, pago, administración, federal, fecha, publica, banco, impuesto, hacienda
- credit, fiscal, payment, administration, federal, date, public, bank, tax, treasury

FREX words:

²³ See: <http://eleconomista.com.mx/sistema-financiero/2014/11/07/cnbv-interviene-ficrea-triangular-recursos>

- credito, tribunal, FICREA, bancaria, juicio, banco, CONDUSEF, ahorro, fiscal, contribuyentes
- credit, tribunal, FICREA (a collapsed credit union under fraud investigation), bank, trial, bank, CONDUSEF (Financial Consumer Protection Agency), saving, financial, contributors (taxpayers)

Top Five Agencies (with proportions of requests in this topic)

SERVICIO DE ADMINISTRACIÓN TRIBUTARIA	0.129
SECRETARÍA DE HACIENDA Y CRÉDITO PÚBLICO	0.125
COMISIÓN NACIONAL BANCARIA Y DE VALORES	0.071
COMISIÓN NACIONAL PARA LA PROTECCIÓN Y DEFENSA DE LOS USUARIOS DE SERVICIOS FINANCIEROS	0.03
INSTITUTO DE SEGURIDAD Y SERVICIOS SOCIALES DE LOS TRABAJADORES DEL ESTADO	0.03

Examples

November 16, 2005: SERVICIO DE ADMINISTRACIÓN TRIBUTARIA

- LISTADO DE EXPEDIENTES VENTILADOS ANTE EL TRIBUNAL FEDERAL DE JUSTICIA FISCAL Y ADMINISTRATIVA CONCLUIDOS A LA FECHA EN QUE SE HAYA DEMANDADO AL SAT EL PAGO DE DAÑOS Y PERJUICIOS CAUSADOS POR SUS SERVIDORES PUBLICOS EN TERMINOS DEL ARTICULO 34 DE LA LEY DEL DEL SAT
- (List of legal suits against the tax ministry pertaining to article 34 of the tax law concluded to date.)

June 3, 2013: SECRETARÍA DE HACIENDA Y CRÉDITO PÚBLICO

- Solicito las cuentas mensuales comprobadas de ingresos que entrega el estado de San Luis Potosí a la Secretaria de Hacienda y Crédito Público (SHCP) del Mes de Diciembre del 2012; Y en relación a los ingresos obtenidos vía convenio de colaboración administrativa en materia fiscal federal del régimen intermedio de actividad empresarial (sección II capítulo II del Título IV de la LISR). Conjuntamente con la información del IVA e IETU y de cualquier otro ingreso federal comprendido en dicho convenio (comprendidos en la cláusula segunda fracciones I a IX y XI de mencionado convenio). Con la información relativa al número de contribuyentes que se dieron de alta baja o cambiaron de régimen del mismo Mes de Diciembre del 2012 y por cada uno de los municipios del Estado de San Luis Potosí.
- (Requests monthly tax accounts that the state of San Luis Potosí reported to the SHCP in December 2012. Specifically references two types of corporate income tax and VAT.)

March 31, 2014: SERVICIO DE ADMINISTRACIÓN TRIBUTARIA

- ¿cuánto dinero ha recaudado el Servicio de Administración Tributaria por concepto de pago de impuestos de las empresas dedicadas a la publicidad exterior: [*names of firms*] en un periodo de tiempo de 2006 a 2013

- (How much tax has been collected from certain advertising firms in the period from 2006 to 2013)

March 20, 2015: COMISIÓN NACIONAL BANCARIA Y DE VALORES

- 1.- La entrega del informe turnado al Comité Técnico del Fondo de protección de sociedades Financieras populares y de Protección a sus ahorradores para que se determinara la Intervención Gerencial la suspensión y posterior Liquidación de Ficrea S.A. de C.V. S.F.P. 2.- Detalle de los pagos efectuados a los ahorradores registrados o a terceros especificando nombre concepto y autorizaciones correspondientes por cualquier concepto de la fecha declaratoria de Liquidación al 31 de marzo de 2015.
- (Requests copy of report on FICREA scandal, and information on any payments made as a result.)

Topic 15: Education

This topic contains a wide variety of requests having to do with education. Requests were grouped together because they contained words pertaining to education and schools, such as education, studies, university, academic, students, teacher. The topic is highly concentrated across agencies, with close to half of requests submitted to the Education Ministry. The other top recipients include universities, the Mexico City Public Schools agency, and CONACYT, Mexico's analogue to the National Science Foundation. However the topic is the fifth-least concentrated geographically. This topic is the fifteenth largest, accounting for 3.7 percent of all requests. The proportion of total requests pertaining to this topic has been relatively stable, with spikes in 2012 and 2013, perhaps in response to the education reform carried out by the newly elected administration.

Individual requests pertain to all issues related to education. Many requests relate to teacher hiring and training, asking about admission into normal schools or the proportion of normal school graduates receiving employment in the federal public education system. Other requests pertain to education spending, accreditation, and the internal processes of the education bureaucracy. Some requests not pertaining to the education sector may have been included because they ask about the qualifications of government personnel, referencing types of degrees and educational attainment.

While it is quite diverse, many requests in this topic are conducive to producing accountability. These could pertain to private concerns over fair treatment in admissions or teacher employment, submitted by the interested parties. Alternatively, many requests are likely submitted by journalists or civil society groups concerned with patronage hiring practices, the electoral engagement of the teachers union and education reform. For instance, in the wake of the 2012-2013 education reform, the federal government has implemented a teacher-evaluation system, which has come under intense scrutiny and been the subject of repeated protests by the dissident faction of the national teachers union, including the threat to obstruct the 2015 election.²⁴

Comparative Measures

- Percent of requests: 3.7% (15/20)
- Herfindahl index across agencies: 0.196 (4/20, fourth-most concentrated)
- Herfindahl index across states: 0.256 (16/20, fifth-least concentrated)

Topwords:

- educación, estudios, escuela, nivel, superior, universidad, curso, academic, tecnología, nacional
- education, studies, school, level, higher (as in higher education), university, course, academic, technology, national

FREX words:

- alumnos, profesores, plantel, escuela, educación, docente, universidad, licenciatura, mujeres, maestría

²⁴ See: <http://www.jornada.unam.mx/2015/06/06/politica/004n2pol>

- students, professors, staff, school, education, teacher, university, bachelor's degree, women, master's degree

Top Five Agencies (with proportions of requests in this topic)

SECRETARÍA DE EDUCACIÓN PÚBLICA	0.431
INSTITUTO POLITÉCNICO NACIONAL	0.069
ADMINISTRACIÓN FEDERAL DE SERVICIOS EDUCATIVOS EN EL DISTRITO FEDERAL (AFSEDF)	0.036
COLEGIO DE BACHILLERES	0.03
CONSEJO NACIONAL DE CIENCIA Y TECNOLOGÍA	0.026

Examples

July 18, 2005: SECRETARÍA DE EDUCACIÓN PÚBLICA

- ¿Cuáles instituciones educativas en la República Mexicana cuentan con estudios de nivel superior vía Internet con validación de la Secretaría de Educación Pública?
- (Which universities in Mexico offer online courses approved by SEP?)

September 9, 2013: INSTITUTO POLITÉCNICO NACIONAL

- Total de aspirantes que presentaron examen para el ciclo escolar 2013-2014/1 para el nivel superior Total de aspirantes que fueron aceptados para el nivel superior para ciclo escolar 2013-2014/1. Del total de aspirantes aceptados para ciclo escolar 2013-2014/1 para el nivel superior cuantos son egresados de los CECyT's del IPN. Total de aspirantes que presentaron examen para el ciclo escolar 2012-2013/1 para el nivel superior Total de aspirantes que fueron aceptados para el nivel superior para ciclo escolar 2012-2013/1. Del total de aspirantes aceptados para ciclo escolar 2012-2013/1 para el nivel superior cuantos son egresados de los CECyT's del IPN.
- (Information on school applicants and who was accepted to the National Politecnic Institute)

March 5, 2015: ADMINISTRACIÓN FEDERAL DE SERVICIOS EDUCATIVOS EN EL DISTRITO FEDERAL (AFSEDF)

- Solicito matriculas correspondientes a la 6a. Generación de la Especialidad de Formación Docente en la Educación Normal de la Escuela Normal Superior de México correspondiente al ciclo escolar 2010-2011. Así como las fechas de titulación de las matriculas correspondientes a la 6a. Generación de la Especialidad de Formación Docente en la Educación Normal de la Escuela Normal Superior de México correspondiente al ciclo escolar 2010-2011.
- (Enrollment in the national teacher training school.)

May 24, 2015: SECRETARÍA DE EDUCACIÓN PÚBLICA

- El índice de reprobación en la Escuela Nacional Preparatoria en las materias de Química Matemáticas e Historia
- (Failure rate at the national preparatory school in chemistry, mathematics, and history.)

Topic 11: Medical Supplies 1: Contracts and Suppliers

This is one of the three topics having to do with medical supplies. Compared to topics 7 and 20, the notable trait of this topic is that many requests have to do with specific documents, such as contracts for medical supplies. (Topic 20 relates more to purchasing, while topic 7 relates to inventories.) Also in contrast to the other two topics, very few requests in this category pertain to medications. This focus is reflected in the top words, such as delivery, provider, code, and request for bids. The FREX top words are almost entirely the names of specific medical supplies (e.g., brushes, catheter, endotracheal tube). This is the single most concentrated topic across agencies, with 72 percent of requests going to the Mexican Social Security Institute. And the topic is close to the median in concentration across states. This is the 16th most requested topic, constituting 2.8 percent of total requests. Requests in this topic have been low in several years, with pronounced spikes in 2007-2009 and 2012-2014, the first three years in each of the two most recent presidential terms.

Individual requests in this topic quite consistently refer to specific documents related to government acquisitions of medical supplies, such as requests for bids, contracts, or documentation of orders. Some requests ask for documents or databases related to a specific provider or order, often mentioning the order number or company name. Others ask for documents recording the results of bidding processes or the analyses that led to contract granting decisions.

This topic has a high potential for accountability. While some requests may have been submitted by medical supplies providers seeking to learn more about the contracting process, it is likely that several requests are made by actors engaged in oversight, seeking to detect underprovision of supplies or favoritism in contract decisions. IMSS acquisitions processes are notoriously vulnerable to corruption in the selection of suppliers as well as waste and underprovision. Two of the (FREX) top words (Anaya and Berrones) likely to refer to a provider who has sold millions of dollars worth of toothbrushes to IMSS clinics and has been the subject of fines from the Ministry of Public Administration for not formalizing contracts.²⁵

Comparative Measures

- Percent of requests: 2.8% (16/20)
- Herfindahl index across agencies: 0.526 (1/20, single most concentrated topic)
- Herfindahl index across states: 0.308 (10/20)

Topwords:

- solicito, documentos, delegación, IMSS, entrega, proveedor, clave, licitación, instituto, bienes
- I request, documents, delegation (outpost of federal agency), IMSS, delivery, provider, code, request for bids, institute, goods

²⁵ See:

<http://compras.imss.gob.mx/?P=imsscomproprod&pd=06018901061101&pr=2009&corder=pr&corderdir=up#detailresultanc> and
http://www01.funcionpublica.gob.mx:8080/SanFicTec/jsp/Ficha_Tecnica/FichaSinTabla.jsp?expe=006410204/2009&tipo=1&persona=3

FREX words:

- cepillos, Distromed, Berrones, endotraqueales, esteril, cateter, reposiciones, dentales, redonda, holiday
- brushes, Distromed, Berrones, endotracheal, sterile, catheter, replacement, dental, round, holiday

Top Five Agencies (with proportions of requests in this topic)

INSTITUTO MEXICANO DEL SEGURO SOCIAL	0.723
INSTITUTO DE SEGURIDAD Y SERVICIOS SOCIALES DE LOS TRABAJADORES DEL ESTADO	0.032
SECRETARÍA DE LA FUNCIÓN PÚBLICA	0.026
SECRETARÍA DE SALUD	0.019
COMISIÓN FEDERAL PARA LA PROTECCIÓN CONTRA RIESGOS SANITARIOS	0.018

Examples

July 3, 2007: INSTITUTO MEXICANO DEL SEGURO SOCIAL

- México DF a 3 de Julio del 2007 A quien corresponda: Solicito una relación de TODAS las órdenes de reposición generadas por el proveedor [*name of supplier*]. en su contrato No [*contract number*] el cual es derivado de la licitación [*public tender number*]. A continuación se presentan otros datos relevantes: * Delegación Coahuila. * Unidad Contratante Del. Coahuila. Esta relación de órdenes de reposición debe incluir la fecha de expedición de la orden la fecha limite de entrega y la fecha real de recepción de bienes. (La respuesta a la solicitud de información No [*request ID#*] ilustra perfectamente lo que se está pidiendo). Favor de presentar la información de tal forma que sea legible. Gracias Atte: El solicitante
- (Requests all orders supplied by a specific company to fulfill a specific contract, including the date or order and date of actual delivery.)

April 18, 2011: INSTITUTO MEXICANO DEL SEGURO SOCIAL

- EN LO QUE VA DEL AÑO 2011 A LA DELEGACIÓN DEL IMSS EN EL ESTADO DE MÉXICO PONIENTE HA LLEVADO A CABO POR LO MENOS UNA ADJUDICACIÓN DIRECTA PARA ADQUIRIR CEPILLOS DENTALES INFANTILES CLAVE [*product code*] EN LA QUE RESULTÓ ADJUDICADO EL PROVEEDOR AURORA ESCAMILLA ZENTENO. EN RELACIÓN A LO ANTERIOR LE SOLICITO A LA DELEGACIÓN DEL IMSS EN EL ESTADO DE MÉXICO PONIENTE LA IMAGEN O EL DOCUMENTO QUE CONSTE QUE LOS CEPILLOS DENTALES INFANTILES CLAVE [*product code*] ESPECIFICAN EN SUS EMPAQUES LA RAZÓN SOCIAL Y RFC DEL DISTRIBUIDOR.
- (Document showing results of public bidding for children's toothbrushes.)

March 15, 2013: INSTITUTO MEXICANO DEL SEGURO SOCIAL

- Le solicito a la Delegación Estatal SINALOA todos los informes analíticos (o documentos equivalentes) emitidos por el laboratorio de control de calidad de [*name of laboratory*] que ha presentado [*name of firm*] al momento de que este proveedor ha hecho sus entregas correspondientes de Tubos endotraqueales ([*product codes*]) para cumplir sus obligaciones contractuales derivadas de la licitación [*public tender number*].
- (Reports from analysis of a specific provider's endotracheal tubes.)

January 1, 2014: INSTITUTO MEXICANO DEL SEGURO SOCIAL

- Solicito a la Delegación Estatal del IMSS en Veracruz Sur la imagen de la etiqueta contenida en los empaques primarios de las sondas tipo NELATON ([*product codes*].) que ha entregado el proveedor [*name of firm*] en cumplimiento de la adjudicación que tuvo en la licitación [*public tender number*].
- (Requests image of package label of a specific type of urinary catheter supplied by a specific supplier for a specific contract.)

Topic 19: Public Procurement 2: Procedures and Documents

This topic is one of three having to do with procurement that is not primarily about medical supplies. Compared with topic 18, which is focused on state-contracted services, this topic tends to include requests for documents and records of procedures, compliance, and fulfillment. The attention to procurement procedures is reflected in top words such as contract, decision, proposal, and procedure. This topic is relatively dispersed across agencies. IMSS received the most requests, but only 19 percent, and a significant number of requests went to the Ministry of Communications and Transportation, the Ministry of Public Administration, and the Electricity Commission. The topic is the seventh most concentrated across states, with 51 percent of requests originating in the Federal District, followed by Mexico State (25 percent) and Morelos (11 percent). All other states had less than 3 percent of requests. This was the 17th largest topic, accounting for 2.3 percent of all requests, although notable spikes occurred in 2004, 2005, and particularly in 2015, where it accounted for over 25 percent of requests in one month.

Individual requests in this topic typically asked for specific documents or records of procurement and contracting processes, often without mentioning the name of the good or service being provided. In several of these cases, the requests mention that the information they are seeking is not available on Compranet, a national database of acquisitions where agencies are required to post requests for bids and decisions. Other requests asked for specific documents related to contracts, such as proof of fulfillment, technical analysis, and scoring of proposals.

Like other topics related to procurement, this topic is likely a mix of requests with accountability and private objectives. Requesters could be engaging in oversight to detect corruption or underprovision or they could be seeking to acquire more information about the procurement process to better tailor their future proposals to future bids. Frequent references to Compranet, and complaints by requesters that they are requesting information that is legally required to be available on this database suggest that the information request process is used a second resource to the limited information proactively made available.

Comparative Measures

- Percent of requests: 2.3% (17/20)
- Herfindahl index across agencies: 0.043 (11/20)
- Herfindahl index across states: 0.337 (7/20)

Topwords:

- licitación, servicios, contrato, técnico, pública, fallo, adquisiciones, propuesta, procedimiento, acta
- request for bids, services, contract, expert, public, decision, acquisitions, proposal, procedure, statement/proceedings

FREX words:

- techo, diferentes, proposiciones, fallo, propuesta, acta, aperture, leds, glosario, adquisiciones
- roof/limit, different, proposal, decision, proposal, statement/proceedings, opening, LEDs, glossary, acquisitions

Top Five Agencies (with proportions of requests in this topic)

INSTITUTO MEXICANO DEL SEGURO SOCIAL	0.184
SECRETARÍA DE COMUNICACIONES Y TRANSPORTES	0.052
INSTITUTO DE SEGURIDAD Y SERVICIOS SOCIALES DE LOS TRABAJADORES DEL ESTADO	0.033
SECRETARÍA DE LA FUNCIÓN PÚBLICA	0.028
COMISIÓN FEDERAL DE ELECTRICIDAD	0.023

Examples

November 11, 2006: INSTITUTO MEXICANO DEL SEGURO SOCIAL

- me gustaria obtener una copia de LOS ANEXOS DE la propuesta economica y de la propuesta tecnica de la LICITACION PUBLICA NACIONAL [REFERENCE NUMBER] LAVADO Y CONFECCION DE CORTINAS Y TAPIZADO DE MOBILIARIO CELEBRADO EN LA CIUDAD DE PUEBLA PUEBLA en el mes de mayo del 2007
- (Requests copies of annexes to the economic proposals and technical proposals for a specific public tender award for curtain cleaning and furniture reupholstery.)

October 26, 2009: INSTITUTO MEXICANO DEL SEGURO SOCIAL

- Solicito a la delegación del IMSS DF Norte los documentos mediante los cuales la empresa [COMPANY NAME] cumplió con lo que señala el inciso J) del numeral 9.1 de las bases de la licitación [REFERENCE NUMBER].
- (Documents through which a specific supplier company complied with a specific provision of a public tender.)

November 25, 2013: HOSPITAL REGIONAL DE ALTA ESPECIALIDAD DE YUCATÁN

- SOLICITUDES DE INFORMACIÓN En materia de Servicio solicito lo siguiente: 1. Contratos de servicio de Promoción y Publicidad Servicio de Publicidad Contratación de Agencia de Publicidad Campaña de Publicidad Agencia de Publicidad Publicidad en medios impresos Publicidad en radio Publicidad en internet Publicidad en librerías Publicidad en medios complementarios así como sus anexos y convenios modificatorios si aplica a partir de 2010 a la fecha. 2. Cuales están vigentes. 3. El monto y vigencia de los mismos. 4. Suficiencia presupuestal de estos. 5. Estudio de Mercado que se realizó para la contratación del servicio. 6. Propuesta Técnica y Económica de los licitantes que llegaron al fallo. 7. Evaluación de Propuestas Técnicas y Económicas. 8. Antecedentes del procedimiento del que derivó el contrato. (Convocatoria acta de junta de aclaraciones acta de presentación y apertura de proposiciones y acta de fallo así como sus respectivos diferimientos.). Lo anterior con fundamento en el artículo 7 fracción IX y XIII de la Ley Federal de Transparencia y Acceso a la Información Pública Gubernamental debido a que información requerida no se encuentra en el sistema COMPRANET.
- (Requests service contracts for any form of advertising, media, or publicity since 2010, as well as annexes and addenda for those contracts, amounts and duration, market studies, technical and economic proposals of the bidders, official evaluations of those proposals, and information on the decision procedure, including meeting minutes.)

January 28, 2014: INSTITUTO MEXICANO DE CINEMATOGRAFÍA

- En materia de Servicio solicito lo siguiente: 1. Contratos de servicio de Mantenimiento y Recarga de Extinguidores Adquisición de Extinguidores Servicio de Recarga de Extinguidores Revisión mensual de extinguidores Adquisición de Portaextinguidores Inspección y Recarga de Extinguidores así como sus anexos y convenios modificatorios si aplica a partir de 2010 a la fecha. 2. Cuales están vigentes. 3. El monto y vigencia de los mismos. 4. Suficiencia presupuestal de estos. 5. Estudio de Mercado que se realizó para la contratación del servicio. 5. Propuesta Técnica y Económica de los licitantes que llegaron al fallo. 6. Evaluación de Propuestas Técnicas y Económicas. 7. Antecedentes del procedimiento del que derivó el contrato. (Convocatoria acta de junta de aclaraciones acta de presentación y apertura de proposiciones y acta de fallo así como sus respectivos diferimientos.). Lo anterior con fundamento en el artículo 7 fracción IX y XIII de la Ley Federal de Transparencia y Acceso a la Información Pública Gubernamental debido a que información requerida no se encuentra en el sistema COMPRANET.
- (Requests contracts for the acquisition, maintenance, and refilling of fire extinguishers since 2010, along with annexes and addenda to those contracts, amounts and duration, market studies conducted, technical and economic proposals of the bidders, evaluations of those proposals, and information on the decision procedure, including meeting minutes.)

Topic 10: Public Procurement 3: Anti-Corruption Campaign

This topic includes elements similar to topics 7, 11, and 20 in its focus on procurement and inventories of supplies and medications in public medical facilities. However many of these requests are similar in that they belong to an anti-corruption campaign, MXSinCorruptos (Mexico without Corruption). Many of the requests include this campaign's twitter handle and motto, and the words in this motto show up as several of the top words. This motto describes the transparency system as a weapon and information requests as the bullets of the people in the battle against corruption. This is the single most concentrated topic across states, with 67 percent of requests coming from Morelos, seemingly the site of the organization directing the MXSinCorruptos campaign. (This is the only topic whose most frequent state is not the Federal District.) The topic is also the third-most concentrated across agencies, with the top three most requested agencies all operating in the health sector (IMSS, the health regulatory agency—COFEPRIS, and the Ministry of Health). The 18th largest topic, this topic accounts for 2 percent of all requests, with a negligible number of requests until 2010, when it picks up and fluctuates between 2 and 8 percent of requests.

While it may seem unusual for such a relatively narrow anti-corruption effort to account for an entire topic of requests, this is a function of the sheer volume of requests filed by this campaign. In fact, according to a report from INAI based on information on user IDs that is not publicly available and so not part of our analysis, a single user ID in the INFOMEX system was responsible for over 25,000 requests from 2007 to 2012.²⁶ This user's requests account for a substantial share of the requests associated with this topic, with the remainder either reflecting requests filed from other user IDs but nonetheless associated with the same anti-corruption campaign, or otherwise requests that are not associated with the campaign, but nonetheless using many similar words.

Typical requests in this topic ask for records of acquisitions and licensing decisions for medications. For instance, dozens of requests were submitting asking about the amount of specific supplies (food, medications, catheters) that were purchased by specific IMSS state delegations or local clinics. Many requests asked for the registro sanitario or analyses of medical supplies and medications, referring to the cases or documents by number.

Given the high prevalence of a requests belonging to a single anti-corruption campaign, it is clear that this topic is geared to generating accountability. This constitutes a striking example of a "super-user"—a civil society activist that relies on the information request system to acquire information in anti-corruption activities. It is clear from the requests that the requesters in this campaign are closely familiar with the transparency requirements in the health sector and have insider knowledge of the specific documents that contain the information of interest.

Comparative Measures

²⁶ "Análisis de demanda de la Administración Pública Federal." 2013.

<http://inicio.ifai.org.mx/Documentos_TP/DGCPA_Presentaci%C3%B3n_Proyecto_Integral13may14.pptx>

- Percent of requests: 2.0% (18/20)
- Herfindahl index across agencies: 0.203 (3/20, third most concentrated)
- Herfindahl index across states: 0.508 (1/20, single most concentrated)

Topwords:

- solicito, aprobada, asamblea, inserciones, información, pública, mexicano, actos, corrupción, siguiente
- request, approved, assembly, insertions, information, public, Mexican, acts, corruption, following

FREX words:

- corruptos, anticorrupción, MXSincorruptos, honestidad, twitter, honremos, contactanos, [email address], recuerda, balas
- corrupt, anticorruption, MXSincorruptos, honesty, twitter, we will honor, contact us, [email address], remember, bullets

Top Five Agencies (with proportions of requests in this topic)

INSTITUTO MEXICANO DEL SEGURO SOCIAL	0.364
COMISIÓN FEDERAL PARA LA PROTECCIÓN CONTRA RIESGOS SANITARIOS	0.23
SECRETARÍA DE SALUD	0.103
SECRETARÍA DE LA FUNCIÓN PÚBLICA	0.078
PROCURADURÍA FEDERAL DEL CONSUMIDOR	0.017

Examples

March 29, 2011: INSTITUTO MEXICANO DEL SEGURO SOCIAL

- Esta Contraloría Social le solicita a la delegación IMSS DF Sur la información documental que consigne el número de las licitaciones públicas de las adjudicaciones directas y/o de las invitaciones a cuando menos tres personas mediante las cuales esta delegación contrató víveres (grupo de suministro 480) en el mes de Febrero de 2011.
PUBLICIDAD DEL SOLICITANTE: EN ESTE 2011 HONREMOS LA LUCHA REVOLUCIONARIA DE 1911 COMBATIENDO SIN TREGUA NI CUARTEL AL CÁNCER DE LA CORRUPCIÓN. EXIJAMOS A NUESTROS SERVIDORES PÚBLICOS EFICACIA HONESTIDAD Y TRANSPARENCIA EN EL DESEMPEÑO DE SU FUNCIÓN PÚBLICA. DEBIDO A QUE TODO MEXICANO TIENE A SU ALCANCE EL FUSIL DE LAS SOLICITUDES DE ACCESO A LA INFORMACIÓN Y LAS BALAS DE LA TRANSPARENCIA ES UNA TRAICIÓN A LA PATRIA EL NO DEFENDER A NUESTRO MÉXICO DE LOS SERVIDORES PÚBLICOS CORRUPTOS QUE LO RAPIÑAN. RECUERDA QUE UN PAÍS EXITOSO Y HONESTO LO FORJAN SUS CIUDADANOS Y ESTA NOBLE TAREA NO SE LA DEBEMOS ENDOSAR A NUESTROS GOBERNANTES. CRUZADA ANTICORRUPCIÓN A.C. TE CONVOCA EN ESTE 2011 A SER PARTE DE LA REVOLUCIÓN DE LA TRANSPARENCIA. SÍGUE Y DENUNCIA ACTOS DE CORRUPCIÓN EN LA CUENTA DE TWITTER @mxsincorruptos DEMOSTREMOS

QUE SOMOS MUCHOS LOS QUE QUEREMOS UN MÉXICO SIN CORRUPOTOS.
(inserción aprobada en la asamblea 2010-12)

- (Requests from a particular IMSS office information on the number of public tenders, direct awards, or invitations to at least three bidders for the supply of a sub-category of food supplies in February 2011. Includes a statement on honoring the revolutionary struggle of 1911 by fighting the cancer of corruption, and a reference to the anti-corruption campaign's Twitter handle.)

December 10, 2013: COMISIÓN FEDERAL PARA LA PROTECCIÓN CONTRA RIESGOS SANITARIOS

- Solicito a la COFEPRIS los documentos que consten el número de visitas de verificación que llevó a cabo esta Comisión para muestrear condones masculinos (o preservativos de hule látex) IMPORTADOS en el año 2010.
- (Requests documents showing the number of inspections COFEPRIS conducted of imported condoms in 2010.)

May 19, 2014: INSTITUTO MEXICANO DEL SEGURO SOCIAL

- SOLICITO A TODAS LAS ÁREAS USUARIAS (COLSULTORIOS QUIROFANOS LABORATORIOS ETC) DEL IMSS DELEGACIÓN CAMPECHE TODA LA INFORMACIÓN DOCUMENTAL QUE CONSTE Y EVIDENCIE LAS QUEJAS PROBLEMAS O DEFECTOS QUE LOS DOCTORES PACIENTES O USUARIOS EN GENERAL HAN ENCONTRADO SUFRIDO O REPORTADO EN RELACIÓN A LOS INSUMOS MÉDICOS PROVEÍDOS POR [company name] EN LO QUE VA DEL AÑO 2014. UN GOBIERNO EFICAZ TRANSPARENTE RESPONSABLE Y CONFIABLE QUE ACTÚE BAJO EL PREDOMINIO DE LA LEY ES LA BASE DE UN DESARROLLO SOSTENIBLE NO EL RESULTADO DE ÉSTE. - KOFI ANNAN SIGUE Y DENUNCIA ACTOS DE CORRUPCIÓN EN LA CUENTA DE TWITTER @mxsincorruptos DEMOSTREMOS QUE SOMOS MUCHOS LOS QUE QUEREMOS UN MÉXICO SIN CORRUPOTOS. (INSERCIÓN APROBADA EN LA ASAMBLEA FEB-14)
- (Requests from the IMSS office in Campeche all documents of complaints, problems, or defects reported concerning medical supplies from a specific supplier company in 2014. Includes a quote from Kofi Annan, and a reference to the anti-corruption campaign's Twitter.)

April 21, 2015: INSTITUTO MEXICANO DEL SEGURO SOCIAL

- Esta Contraloría Ciudadana le Solicita a la UMAE 71 TORREON COAHUILA todos los certificados analíticos certificados de calidad o documentos equivalentes que ha entregado el proveedor [company name] al momento de hacer sus entregas de guante de polietileno cuyas claves son 060.456.0037.11.01 y 060.456.0045.11.01 los cuales le fueron asignados en el procedimiento de contratación [reference number].
PUBLICIDAD DELSOLICITANTE: PERTENEZCO A UNA GENERACIÓN QUE QUISO CAMBIAR EL MUNDO. FUI APLASTADO DERROTADO PULVERIZADO PERO SIGO SOÑANDO QUE VALE LA PENA LUCHAR PARA QUE LA GENTE PUEDA VIVIR UN POCO MEJOR Y CON UN MAYOR SENTIDO DE IGUALDAD - JOSÉ MÚJICA SIGUE Y DENUNCIA ACTOS DE CORRUPCIÓN EN LA CUENTA

DETWITTER @mxsincorruptos DEMOSTREMOS QUE SOMOS MUCHOS LOS QUE QUEREMOS UN MÉXICO SIN CORRUPTOS. (INSERCIÓN APROBADA EN LA ASAMBLEA MAR-15)

- (Requests quality certificates for polyethylene gloves supplied to a specific hospital by a specific provider. Includes an inspirational quote from the former President of Uruguay, and a reference to the anti-corruption campaign's Twitter.)

Topic 20: Medical Supplies 2: Purchases and Spending

This topic is one of the three topics having to do with medical supplies. Compared with topic 11, which pertains primarily to bids and contract documents, this topic pertains more to actual spending and acquisitions of medications and other supplies. Most requests refer to the process of selecting providers and purchasing, as reflected in top words such as purchase, price, products, and storage. The sixth most concentrated topic across agencies, the vast majority of requests were sent to ministries that administer public health programs; IMSS, the Ministry of Health and ISSSTE were the three ministries that each received over 10 percent of requests. This is the second-most concentrated topic across states, with 65 percent of requests coming from the Federal District and 24 percent from Mexico State. The second-smallest topic, this topic accounts for 1.2 percent of all requests with general low request numbers and modest spikes in the post-2009 period.

Requests in this topic often ask for aggregate data or lists of medications and supplies purchased by government ministries. Some requests also asked about personnel at health clinics. Rather than referring to specific documents or contractors, the most common format of requests asked about the total amounts of supplies acquired, budgets spent, and other aggregate data for specific facilities. Many requests were submitted in the same format, with a lengthy preamble referencing the text of the transparency law.

This topic is likely a mix of requests with accountability and private objectives. As discussed in topics 7, 10, and 11, these requests seek information that could be useful in uncovering corruption or underprovisioning of services in public health facilities. Given that many requests were submitted in the same format and make explicit reference to the transparency law, it is clear that this topic includes many requests from “elite” actors who are aware of the type of information that exists and are interested in acquiring spreadsheets to analyze medical purchasing. This type of information also has clear private purposes, however, such as in conducting market research for producers of medical supplies.

Comparative Measures

- Percent of requests: 1.2% (19/20)
- Herfindahl index across agencies: 0.083 (6/20, sixth most concentrated)
- Herfindahl index across states: 0.479 (2/20, second most concentrated)

Topwords:

- compra, favor, número, información, medicamento, precio, descripciones, importe, presente, productos
- purchase, please, number, information, medication, price, descriptions, price, current, products

FREX words:

- medicamento, compra, realizer, diferencial, precio, restringida, almacen, domicilio, párrafo, vendió
- medication, purchase, to carry out, differentiating, price, restricted, storage, address, paragraph, sold

Top Five Agencies (with proportions of requests in this topic)

INSTITUTO MEXICANO DEL SEGURO SOCIAL	0.209
SECRETARÍA DE SALUD	0.136
INSTITUTO DE SEGURIDAD Y SERVICIOS SOCIALES DE LOS TRABAJADORES DEL ESTADO	0.112
PETRÓLEOS MEXICANOS	0.062
SECRETARÍA DE LA DEFENSA NACIONAL	0.033

Examples

October 7, 2008: HOSPITAL REGIONAL DE ALTA ESPECIALIDAD DE YUCATÁN

- Adquisición y compra de productos auxiliares para la salud (material de curación material de laboratorio y radiológico) en formato Excel hechas por el Hospital RAE de Yucatán para el primer segundo y tercer trimestre de 2008 indicando por cada trimestre la adquisición de manera separada
- (Requests an Excel spreadsheet of all purchases and acquisitions of ancillary health products in the first three quarters of 2008.)

February 13, 2009: INSTITUTO MEXICANO DEL SEGURO SOCIAL

- Conforme a la relación que se anexa indique por favor para cada clave de producto y proveedor de la siguiente Adjudicación Directa [*contract number*] el nombre del fabricante según la notificación de apoyo recibida por su institución en el 2008. La información requerida es por clave de cada producto según el cuadro básico de medicamentos su descripción el nombre del proveedor del mismo el nombre del fabricante que apoya al proveedor cuando se trata de un distribuidor la modalidad de compra la cantidad adquirida precio unitario importe cantidad máxima cantidad mínima así como indicar la cantidad que el fabricante aporta por cada clave número de licitación y número de contrato para cada delegación y/o UMAE (Unidad Médica de Alta Especialidad).
- (Detailed information on the manufacturer, products, and purchases for each product in the attached file for a direct award contract.)

January 1, 2012: INSTITUTO MEXICANO DEL SEGURO SOCIAL

- Con fundamento en lo previsto en los Artículos 4 7 (párrafo ultimo) 9 13 17 18 19 40 (párrafo ultimo) 43 y 63 (inciso VI) de la Ley General de Transparencia y acceso a la información Pública Gubernamental y con fundamento en el título segundo de la misma donde se aclara a la ciudadanía las responsabilidades de las unidades de enlace y considerando que en los términos del Capítulo III la presente solicitud no está abarcando ninguna información confidencial y la información debe de entregarse en los tiempos establecidos en el artículo 44 se expide la presente solicitud. Favor de indicar la Compra Real de todos los MATERIALES DE CURACION del INSTITUTO MEXICANO DEL SEGURO SOCIAL en el periodo OCTUBRE - DICIEMBRE DEL 2011. Datos requeridos únicamente: Clave Cuadro Básico Descripción completa y clara de los materiales de curación Número de piezas compradas Precio por pieza Importe

Proveedor que vendió el material de curación Tipo de Compra Número de Licitación ó Adjudicación Directa ó Invitación Restringida según corresponda Número de Contrato o Factura Almacén o Unidad Médica que realizó la compra.

- (Record of all medical materials purchased by IMSS between October and December 2011, including description of materials, number of units, price, provider, type of purchase, number of contract, and unit making the purchase.)

October 1, 2014: SECRETARÍA DE SALUD

- Favor de indicar la Compra real de todos los Medicamentos y Productos Farmacéuticos de CENSIA (CENTRO NACIONAL PARA LA SALUD DE LA INFANCIA Y LA ADOLESCENCIA) en el periodo del 01 de Julio al 30 de Septiembre del 2014. Datos requeridos: Clave Cuadro Básico Descripción completa y clara del medicamento presentación del producto (genérico patente o innovador) origen del producto nivel de atención en el que se utiliza el medicamento proveedor ganador fabricante del medicamento marca del medicamento (en el caso de que corresponda) fuentes de abasto (100% 60%/40% etc. solo en caso de aplicar) Número de piezas asignadas por medicamento Precio asignado por Pieza Importe Tipo de adquisición gubernamental en el que se convocó (Licitación Adjudicación Directa Invitación Restringida) según corresponda Numero de la adquisición gubernamental convocada (Licitación Adjudicación Directa Invitación Restringida) según corresponda indicar el criterio de asignación utilizado para la realización de la adquisición gubernamental (criterio binario puntos y porcentajes etc.) Almacén o Unidad Médica (Hospital Clínica o Consultorio) en donde se entregara el producto.
- (Information about medications and pharmaceutical supplies purchased by a children's health center with several details about each purchase: price, type of purchase, acquisition number, etc.)

Topic 7: Medical Supplies 3: Inventories

This topic has to do with inventories in public health facilities, referring mostly to medications. Rather than asking about contracts or purchasing, the focus of topics 10, 11, and 20, this topic is focused on the actual stores of supplies. The second-most concentrated topic across agencies, 65 percent of requests were submitted to IMSS. This is also the fourth most concentrated topic across states, with 49 percent of requests from the Federal District and 40 percent from Mexico State. The smallest of all topics, the requests in this category are concentrated in the period from 2009 to 2014.

Individual requests in this topic tended to ask about the inventory of certain products, especially medications, at specific clinics. Many requests were in the same format, only varying the name of the clinic or medication. Other requests asked for aggregate data on the number of medical procedures carried out or the budgets allocated for specific products.

Requests in this topic may be oriented to producing accountability, as information about stocks of supplies and medications would be useful in campaigns to uncover underprovision in public health services. However, other requests about aggregate services and supplies could be useful to journalists or industry analysts researching medical services in Mexico, or conducting market research for drug and equipment producers .

Comparative Measures

- Percent of requests: 0.6% (20/20)
- Herfindahl index across agencies: 0.430 (2/20, second most concentrated)
- Herfindahl index across states: 0.409 (4/20, fourth most concentrated)

Topwords:

- cada, mes, clave, existentes, inventario, cuadro, salud, básico, entradas, final
- each, month, code, existing, inventory, box, health, basic, entered, final

FREX words:

- tabletas, parche, nicotina, ácido, acetilsalicílico, paracetamol, acetaminophen, centímetros, masticable, solubles
- tablets, patch, nicotine, acid, aspirin, acetaminophen, acetaminophen, centimeters, chewable, dissolvable

Top Five Agencies (with proportions of requests in this topic)

INSTITUTO MEXICANO DEL SEGURO SOCIAL	0.654
INSTITUTO DE SEGURIDAD Y SERVICIOS SOCIALES DE LOS TRABAJADORES DEL ESTADO	0.046
SECRETARÍA DE SALUD	0.019
COMISIÓN NACIONAL DEL AGUA	0.012
PETRÓLEOS MEXICANOS	0.012

Examples

March 8, 2011: INSTITUTO MEXICANO DEL SEGURO SOCIAL

- FAVOR DE INDICAR EL INVENTARIO INICIAL ENTRADAS SALIDAS INVENTARIO FINAL EXISTENCIAS MAXIMAS Y EXISTENCIAS MINIMAS DE PIEZAS POR CADA MES (JULIO DEL 2011) DE TODOS LOS MEDICAMENTOS LÁCTEOS Y PRODUCTOS FARMACÉUTICOS PARA EL ALMACEN DELEGACIONAL COLIMA. DATOS REQUERIDOS ÚNICAMENTE: EL MES DE JULIO DEL 2011 NOMBRE DE LA UNIDAD MEDICA (A DONDE DESPLAZA ESTAS CLAVES SI ES EL CASO) CLAVE DEL MEDICAMENTO DESCRIPCIÓN COMPLETA Y CLARA DE LA CLAVE NÚMERO DE PIEZAS POR CLAVE DEL CUADRO BASICO DE EXISTENCIAS INICIALES POR CADA MES (INVENTARIO INICIAL) ENTRADAS DE PIEZAS POR CADA CLAVE DEL CUADRO BASICO DE CADA MES SALIDAS DE PIEZAS POR CADA CLAVE DE CUADRO BASICO DE CADA MES NÚMERO DE PIEZAS POR CLAVE DE EXISTENCIAS FINALES AL CIERRE DE CADA MES (INVENTARIO FINAL) EXISTENCIAS MAXIMAS Y MINIMAS DE PIEZAS POR CADA CLAVE DEL CUADRO BASICO DE MEDICAMENTOS Y CONSUMO PROMEDIO MENSUAL DE JULIO DEL 2011 PARA ESTE ALMACEN.
- (Initial inventory, incoming, departing, and final inventory for July 2011 of all milk products and pharmaceutical products for the Colima IMSS delegation, including name of medical unit, product codes and other information.)

May 2, 2013: : INSTITUTO MEXICANO DEL SEGURO SOCIAL

- Solicito atentamente me indique el número de intervenciones quirúrgicas que fueron realizadas anualmente en las Unidades Médicas de la Delegación que a continuación se indica para cada una de las cirugías señaladas a continuación (señalar cada intervención por cada tipo de cirugía a señalar) para el año 2012 con el mayor detalle posible:
QUINTANA ROO HGSMF 2 COZUMEL HGR 17 CANCUN HGZMF 1 CHETUMAL HGZ 3 CANCUN HGZ 18 PLAYA DEL CARMEN HGP 7 PUERTO JUÁREZ
1. Cesárea y extracción de feto 2. Excisión de laringe 3. Excisión de pulmón y de bronquio 4. Extracción de cuerpo extraño y cálculo no operatorio 5. Extracción y restauración de dientes 6. Incisión y excisión de cráneo cerebro y meninges cerebrales 7. Incisión y excisión de estructuras de la articulación 8. Incisión y extirpación del estómago 9. Incisión excisión y división de otros huesos 10. Incisión excisión y oclusión de vasos 11. Incisión extirpación y anastomosis de intestino 12. Operaciones de la córnea 13. Operaciones de reparación y plásticas sobre estructuras de la articulación 14. Operaciones reconstructivas sobre oído medio 15. Operaciones sobre amígdalas y adenoides 16. Operaciones sobre ano 17. Operaciones sobre aparato lagrimal 18. Operaciones sobre apéndice 19. Operaciones sobre cordón espermático epidídimo y conducto deferente 20. Operaciones sobre el cristalino 21. Operaciones sobre el cuello uterino 22. Operaciones sobre el oído externo Incluye: operaciones sobre conducto auditivo externo piel y cartílago de meato pabellón 23. Operaciones sobre el pene Incluye: operaciones sobre: cuerpos cavernosos glánde del pene prepucio 24. Operaciones sobre el riñón 25. Operaciones sobre el sistema linfático 26. Operaciones sobre escroto y túnica vaginal 27. Operaciones sobre esófago 28. Operaciones sobre faringe Incluye: operaciones sobre

bolsa faríngea hipofaringe nasofaringe orofaringe seno piriforme 29. Operaciones sobre glándulas y conductos salivales 30. Operaciones sobre hígado 31. Operaciones sobre huesos y articulaciones faciales 32. Operaciones sobre iris cuerpo ciliar esclerótica y cámara anterior del ojo 33. Operaciones sobre la conjuntiva 34. Operaciones sobre la lengua 35. Operaciones sobre la mama 36. Operaciones sobre la nariz 37. Operaciones sobre la pared torácica pleura mediastino y diafragma 38. Operaciones sobre la vagina y fondo de saco 39. Operaciones sobre la vulva y el perineo 40. Operaciones sobre las glándulas tiroides y paratiroides 41. Operaciones sobre las trompas de Falopio 42. Operaciones sobre los ovarios 43. Operaciones sobre médula espinal y estructuras del conducto espinal 44. Operaciones sobre médula ósea y bazo 45. Operaciones sobre músculo tendón y fascia de mano 46. Operaciones sobre músculo tendón fascia y bolsa sinovial salvo de la mano 47. Operaciones sobre músculos extraoculares 48. Operaciones sobre nervios craneales y periféricos 49. Operaciones sobre nervios o ganglios simpáticos 50. Operaciones sobre órbita y globo ocular 51. Operaciones sobre otras glándulas endocrinas 52. Operaciones sobre páncreas 53. Operaciones sobre párpados 54. Operaciones sobre piel y tejido subcutáneo 55. Operaciones sobre próstata y vesículas seminales 56. Operaciones sobre recto rectosigmoide y tejido perirrectal 57. Operaciones sobre retina coroides cuerpo vítreo y cámara posterior del ojo 58. Operaciones sobre senos nasales (paranasales) 59. Operaciones sobre testículos 60. Operaciones sobre uréter

- (Number of surgeries carried out in several different IMSS facilities in the state of Quintana Roo in 2012, including a list of 60 different types of surgery, in as much detail as possible.)

May 9, 2013: Secretaría de Marina

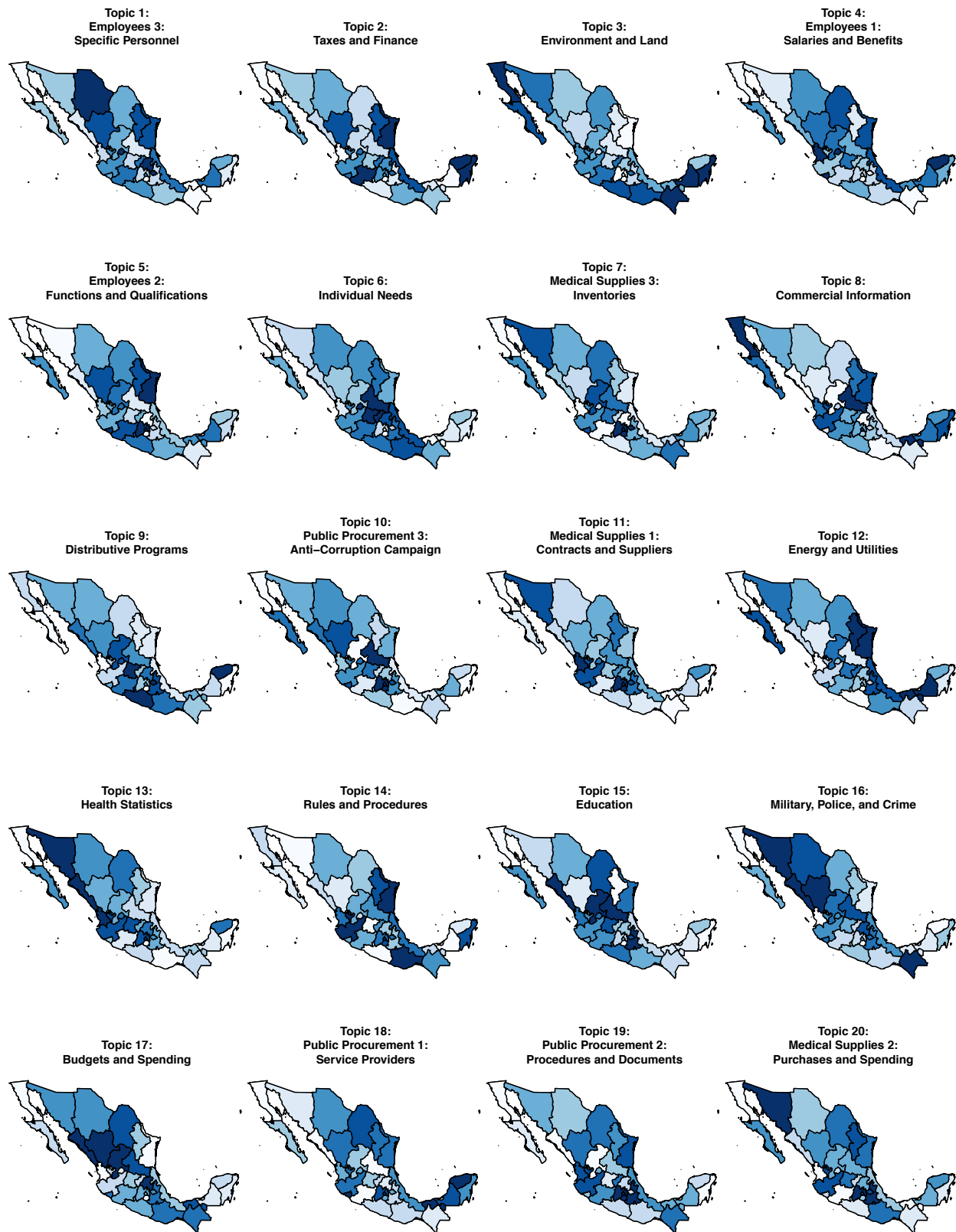
- Deseo conocer toda la información correspondiente a la (s) Delegación Federal Hospital Escuela representación oficina centro de trabajo unidad administrativa o cualquier otra forma de organización o representatividad que se tenga en el estado de YUCATAN relativa al consumo adquisición y proveedores de agua purificada durante el ejercicio 2012 y 2013; desglosado por mes y por Dirección o Jefatura Subdirección o Subdelegación Coordinación Jefatura centro de trabajo unidad administrativa oficina y/o cualquier otra organización administrativa que se tenga. Los datos que respetuosamente solicito son los siguientes: 1.- Nombre del o los proveedores de agua purificada. 2.- Montos pagados por concepto de agua purificada. 3.- Cuantos garrafones de 20 litros se consumen. 4.- Cuantas botellas de agua purificada de otras presentaciones se consumen. 5.- Consumo total en litros por mes. 6.- Cantidad de empleados. 7.- Número de dispensadores. (Por dispensador se refiere a todos los equipos que se tengan para suministrar agua purificada) 8.- Ubicación de cada uno de los dispensadores precisando su tipo y capacidad es decir si llevan garrafones de 20 litros o alguna otra capacidad y si suministran agua fría y caliente o solo al tiempo. Para mejor proveer se anexan dos tablas ejemplificativas de la información que se solicita: Favor de utilizar cuantas filas sean necesarias para que la información este claramente desagregada por unidad administrativa y concepto (proveedor monto concepto desglosado presentaciones consumo No. de dispensadores ubicación tipo y capacidad).
- (Information related to acquisition of purified water in 2012 and 2013 by hospitals operated by the Navy in the state of Yucatan, disaggregated by month and unit and

including a list of details, such as the name of the providers, prices, location of dispersers, etc.)

April 10, 2011: INSTITUTO DE SEGURIDAD Y SERVICIOS SOCIALES DE LOS TRABAJADORES DEL ESTADO

- PARA LA SUBDIRECCIÓN DE ABASTO DE INSUMOS MÉDICOS. FAVOR DE INDICAR EL INVENTARIO INICIAL ENTRADAS SALIDAS INVENTARIO FINAL EXISTENCIAS MAXIMAS Y EXISTENCIAS MINIMAS DE FORMA MENSUAL PARA EL MES DE SEPTIEMBRE DEL 2011 DE TODOS LOS MEDICAMENTOS LÁCTEOS Y PRODUCTOS FARMACÉUTICOS. DATOS REQUERIDOS ÚNICAMENTE: INFORMACION UNICAMENTE DEL MES DE SEPTIEMBRE DEL AÑO 2011 INDICANDO ALMACÉN REGIONAL O UNIDAD MÉDICA CLAVE DEL MEDICAMENTO DESCRIPCIÓN CLARA DE LA CLAVE NÚMERO DE PIEZAS POR INVENTARIO INICIAL ENTRADAS SALIDAS INVENTARIO FINAL EXISTENCIAS MAXIMAS Y EXISTENCIAS MINIMAS AL CIERRE DEL MES DE SEPTIEMBRE DEL 2011 DE CADA UNIDAD MÉDICA DEL ISSSTE.
- (Initial inventory, incoming, departing, and final inventory for September 2011 of all milk products and pharmaceutical products for all medical units of ISSSTE, including name of medical unit, product codes and other information.)

Maps of Topic Distributions



Unstemmed Topwords in Spanish

Ten highest probability words for each topic:

Topic	Top 10 Words	Labels
1	mexicano licitación josé general director colonia delegación copia colonia luis	Employees 3: Personnel
2	credito fiscal pago administración federal fecha publica banco impuesto hacienda	Taxes and Finance
3	municipio ubicado zona proyecto agua solicito construcciones san ambiental federal	Environment and Land
4	trabajo solicito año personal sueldo social seguridad salario pago base	Employees 1: Salaries/Benefits
5	publica administración nombre direcciones general area servidores organismo funcionarios cargo	Employees 2: Functions/Qualif.
6	saber información puede gracias hacer caso servicios alguna existencias debe	Individual Needs
7	cada mes clave existentes inventario cuadro salud basico entradas final	Medical 3: Inventories
8	solicito copia documentos fecha información solicitud numero oficio registro certificada	Commercial Information
9	programa proyecto desarrollo nacional apoyo social realizar recursos cual acciones	Distributive Programs
10	solicito aprobada asamblea inserciones información publica mexicano actos corrupción siguiente	Procurement 3: Anti-Corruption
11	solicito documentos delegación imss entrega proveedor clave licitación instituto bienes	Medical 1: Contracts/Suppliers
12	pemex mexicano servicios año empresa operación tipo electrónico transporte producciones	Energy and Utilities
13	medicamento año numero hospital solicito salud especial edad información nacional	Health Statistics
14	ley publica federal información articulo acceso transparente unidad gubernamental conforme	Rules and Procedures
15	educación estudios escuela nivel superior universidad curso academico tecnología nacional	Education
16	cuanto año numero solicito información personal cual federal fecha caso	Military, Police, and Crime
17	cuanto año presupuesto gastos monto cada destino solicito información parte	Budgets and Spending
18	contrato servicios empresa monto solicito nombre dependientes licitación fecha seguridad	Procurement 1: Service Providers
19	licitación servicios contrato tecnico publica fallo adquisiciones propuesta procedimiento acta	Procurement 2: Procedures/Docs
20	compra favor numero información medicamento precio descripciones importe presente productos	Medical 2: Purchases/Spending

Ten most frequent and exclusive words for each topic:

Topic	Top 10 Words	Labels
1	perez huasteca ortiz vargas javier castro primer hector muñoz aguilar	Employees 3: Personnel
2	credito tribunal ficrea bancaria juicio banco condusef ahorro fiscal contribuyentes	Taxes and Finance
3	ejido predio terrenos sueldo forestal superficie tramite semarnat minera ambiental	Environment and Land
4	jubilación pensiones sindicato tabulador sueldo salario trabajo percepciones salarial nomina	Employees 1: Salaries/Benefits
5	villagomez organismo servidores internacional funcionarios vacantes funcionarios puesto carrera cargo	Employees 2: Functions/Qualif.
6	the puede hacer espera antemano ayuda quiero saber acudir quisiera	Individual Needs
7	tabletas parche nicotina acido acetilsalicilico paracetamol acetaminofen centimetros masticable solubles	Medical 3: Inventories
8	simple certificada oficio expediente copia folio permiso expediente sanitario sorteos	Commercial Information
9	estrategia programa rural indigenas desarrollo prevenciones fomento metros poblaciones cooperación	Distributive Programs
10	corruptos anticorrupción mxsincorruptos honestidad twitter honremos contactanos [email] recuerda balas	Procurement 3: Anti-Corruption
11	cepillos distromed berrones endotraqueales esteril cateter reposiciones dentales redonda holiday	Medical 1: Contracts/Suppliers
12	alumbrado gastos tarifa energia diesel pemex cfe telecomunicaciones refinación exportación	Energy and Utilities
13	pacientes hospitalaria recetas morbilidad medicina hospital tratamiento sexo enfermedades medicamento	Health Statistics
14	reglamento dispuesto ley articulo fracciones acceso articulo disposiciones constitución conforme	Rules and Procedures
15	alumnos profesores plantel escuela educación docente universidad licenciatura mujeres maestría	Education
16	policia detenidos decomisadas militar narcotrafico victimas delitos migratoria crimen roberto	Military, Police, and Crime
17	gastos viajes viaticos presupuesto destino donativos biblioteca presidente dinero spots	Budgets and Spending
18	pef prestación maniobras fletes contrato celebrado poliza paqueter empresa formalizados	Procurement 1: Service Providers
19	techo diferentes proposiciones fallo propuesta acta apertura leds glosario adquisiciones	Procurement 2: Procedures/Docs
20	medicamento compra realizar diferencial precio restringida almacen domicilio parrafo vendió	Medical 2: Purchases/Spending

Bibliography

- Ahonen, P. (2016). Digital public-administration research drawing from Bayesian inspiration: latent trait scaling and topic modeling examination of budgetary legislation in thirteen countries. *Administrative Culture*. 17(1): 47-70.
- Bagozzi, B. E. (2015). The multifaceted nature of global climate change negotiations. *Review of International Organizations*. 10(4): 439-464.
- Bagozzi, B. E. and Berliner, D. (2017). The politics of scrutiny in human rights monitoring: Evidence from structural topic models of u.s. state department human rights reports. *Political Science Research and Methods*.
- Bagozzi, B. E. and Schrodtt, P. A. (2012). The dimensionality of political news reports. Paper presented at the EPSA Annual Convention.
- Baille, M., Carman, M. J., and Crestani, F. (2009). A topic-based measure of resource description quality for distributed information retrieval. *Advances in Informational Retrieval*.
- Barberá, P., Bonneau, R., Jost, J. T., Nagler, J., and Tucker, J. (2014). Is there anybody out there? The effects of legislators' communication with their constituents. Paper presented at the ISA Annual Convention.
- Blei, D. M., Ng, A., and Jordan, M. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993-1022.
- Grimmer, J. (2012). A bayesian hierarchical topic model for political texts: Measuring expressed agendas in senate press releases. *Political Analysis*, 18(1):1-35.
- Grimmer, J. and Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3):267-297.
- Grün, B. and Hornik, K. (2011). topicmodels: An R package for fitting topic models. *Journal of Statistical Software*, 40(13):1-30.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). The Elements of Statistical Learning. Springer.
- Hopkins, D. J. and King, G. (2010). A method for automated nonparametric content analysis for social science. *American Journal of Political Science*, 54(1): 229-247.
- INAI (2015). Sistema infomex gobierno federal. URL: <https://www.infomex.org.mx/gobiernofederal/homeOpenData.action>. Accessed on: 7/20/2015.
- Liénoú, M., Maître, H., and Datcu, M. (2010). Semantic annotation of satellite images using latent dirichlet allocation. *IEEE Geoscience and Remote Sensing Letters*, 7(1):28-32.

Priva, U C. and Austerweil, J. L. (2015). Analyzing the history of cognition using topic models. *Cognition*. 134: 4-9.

Quinn, K. M., Monroe, B. L., Colaresi, M., Crespín, M. H., and Radev, D. R. (2010). How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science*, 54(1):209-228.

Qumsiyeh, R. and Ng, Y.-K. (2011). Readaid: A robust and fully-automated readability assessment tool. *23rd IEEE International Conference on Tools with Artificial Intelligence*.

Rice, D. (2012). Measuring the issue content of supreme court opinions through probabilistic topic models. Presented at the 2012 Midwest Political Science Association Conference, Chicago.

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gdarian, S., Albertson, B., and Rand, D. (2014). Structural topic models for open ended responses. *American Journal of Political Science*, 58(4):172-178.

Yano, T., Cohen W. W., and Smith, N. A. (2009). Predicting response to political blog posts with topic models. *Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the ACL*: 477-485.