



Large-N bill positions data from MapLight.org: What can we learn from interest groups' publicly observable legislative positions?

Geoffrey M. Lorenz¹  · Alexander C. Furnas² · Jesse M. Crosson³

Published online: 22 April 2020
© Springer Nature Limited 2020

Abstract

The transparency organization MapLight records instances of organizations taking positions for and against legislation in Congress. The dataset comprises some 130,000 such positions taken on thousands of bills between the 109th and 115th Congresses (2005–2018). The depth and breadth of these data potentially give them wide applicability for answering questions about interest group behavior and influence as well as legislative politics more broadly. However, the coverage and content of the data are affected by aspects of MapLight's research process. This article introduces the MapLight dataset and its potential uses, examines issues related to sampling and other aspects of MapLight's research process, and explains how scholars can address these to make appropriate use of the data.

Keywords MapLight · Interest groups · Legislative politics · Political economy · Lobbying · Bill positions · Congress

Interest groups routinely lobby for and against legislation in the U.S. Congress. These efforts may influence a bill's content as well as its likelihood of advancing through the legislative process and into law. For decades, scholarship on interest groups has been hampered by the difficulty of assessing groups' positions on

✉ Geoffrey M. Lorenz
gmlorenz@unl.edu

Alexander C. Furnas
zfurnas@umich.edu

Jesse M. Crosson
jcrosson9@gmail.com

¹ Department of Political Science, University of Nebraska-Lincoln, Lincoln, USA

² Department of Political Science, University of Michigan, Ann Arbor, USA

³ Trinity University and Faculty Fellow, Center for the Study of Democratic Politics, Princeton University, Princeton, USA



particular issues or bills. Interest group and lobbying scholars have made frequent and extensive use of federal lobbying registration and activity disclosures mandated by the Lobbying Disclosure Act (LDA) of 1995 (e.g., Baumgartner and Leech 2001; LaPira and Thomas 2017; Bertrand et al. 2011; Vidal et al. 2012; Furnas et al. 2019; Grossmann and Pyle 2013; You 2017; Kim 2017, 2018). But while disclosures have improved over time, they remain incomplete as a source of information about interest groups' specific objectives and consequent influence, and therefore are of limited use for analyses of legislative politics more broadly. In particular, although LDA reports now often mention specific bill names and bill numbers, they rarely contain information about the registering organization's valence toward the bill in question—i.e., whether the group supported or opposed the legislation or its major policy components. Without this information, it is difficult to assess whether bill passage or failure counts as a “success” for the interest group, limiting an analyst's ability to assess a group's influence over specific legislative outcomes.

The nonprofit, nonpartisan transparency organization MapLight offers a potential solution to this problem. As part of its ongoing research into the role of interest groups and money in American politics, MapLight records instances of organizations taking positions on congressional bills. As of May 2019, they have recorded over 130,000 expressions of support and opposition made by over 16,000 unique organizations with respect to nearly 10,000 unique bills introduced between the 109th and 115th Congresses (2005–2018). These positions come from public statements (e.g., on organizations' websites), open letters to Congress, news stories, congressional hearing testimony, and other publicly observable sources. Recorded positions are aggregated and released to the public as a database via MapLight's Bill Positions application programming interface (API). This database has the potential to inform case studies on individual issues or bills and to facilitate large-n, generalizable research on interest group behavior and influence. Indeed, usage of the complete MapLight dataset has expanded in recent years (Lorenz 2020; Crosson et al. 2019; Fagan et al. 2019). However, there is yet no systematic assessment of the suitability of MapLight data for making general inferences about interest group activity and its influence on legislative agendas, processes, and outcomes.

Here, we examine the MapLight Bill Positions data for its potential utility to political science research and provide recommendations concerning its use. We proceed as follows. First, we introduce the reader to examples of existing research employing the MapLight data to study interest group behavior and influence, as well as other related topics such as legislative politics and political economy. Second, we describe what is included in the data as MapLight releases it. Third, we examine how MapLight selects bills for research, and discuss the implications of their selection process for the kinds of descriptive and causal inferences that the data might be used to make. Fourth, we broaden our discussion of the strengths and potential limitations of the MapLight data, including comparisons to alternative large-n sources of data on interest group activity. Fifth, we provide guidance on two methods for acquiring the MapLight dataset. Finally, for scholars who are interested in position-taking data but find the MapLight dataset inappropriate for their purposes, we point toward alternative sources of position-taking data.



Example uses of MapLight data

In this section, we examine several different uses of MapLight's position-taking data. In many cases, the works examined here are forthcoming, further suggesting the yet-untapped richness of the MapLight data for empirical analyses. Emerging uses of the MapLight data fall into three general categories. The first set of studies use MapLight data to investigate individual bill or issue area dynamics. The second set use the entire MapLight dataset to make general claims on the role of interest groups in legislative politics across bills and issue areas. Finally, the third set use the MapLight data to generate summary measures about organizations, legislators, and bills. While not yet numerous, these examples demonstrate a breadth of different applications and may in turn inform readers' intuitions about whether the MapLight data might apply to their own research questions.

The most common use of MapLight data to date has been to focus on particular policy areas. Galantucci (2015) examines lobbying on two bills related to currency manipulation, finding that members of Congress receiving campaign contributions from industries economically dependent on China often withheld support for legislation meant to combat its exchange rate setting practices. He also finds that members representing districts dependent on exports to China were more likely to withhold their support for such legislation. Broz (2016), examining a similar topic (including one of the same bills), highlights the role of global supply chains and exchange rate pass-throughs in legislator support for these bills. In addition to examining the position-taking of different industries within a given bill, MapLight data are useful for examining a single industry across many bills. Moore et al. (2013) examine the relative influence of the automotive industry's campaign contributions and district presence (e.g., factories) on legislators' roll-call votes, using the MapLight data to identify "pro-automotive industry" positions on approximately 250 bills. Using the MapLight data and connecting it to information about bill contents, they are able to show that both contributions and district presence appear to influence roll-call votes on salient issues, but only when the bills involved were narrowly tailored to automaker interests (e.g., the automotive bailout and the "Cash for Clunkers" program). As these examples illustrate, MapLight data can provide rich information about the organizations and interests advocating within particular policy areas, or the ability of particular industries to attain their preferences. All of these applications require an organization's valence (support or oppose) on individual bills; MapLight data are uniquely suited to meeting this requirement.

Beyond MapLight data's usefulness for examining individual bills or industries, an important feature of the data is its ability to facilitate claims that are generalizable across issue areas. Lorenz (2020) analyzes the complete MapLight dataset (over 4700 bills at the time) to explain lobbying coalition influence on the advancement of bills through committee markup and reporting. He finds that lobbying coalitions comprised of a diverse set of industries, social causes, and other interests are more strongly associated with committee consideration than other types of lobbying coalitions. Furthermore, he shows that the association between



diverse lobbying coalitions and legislative agenda-setting is conditional on institutional partisan alignments and the degree to which agenda-setters value information about a bill's downstream viability. Turning from interest group influence to interest group behavior, Fagan et al. (2019) show that groups more strongly aligned with one of the major parties are frequently drawn into policy conflicts across a wider range of issue areas. Finally, using a special version of the MapLight dataset that captures the timing of position-taking actions relative to bill advancement, Holyoke (2019) shows that groups tend to take positions at times that are convenient for their allies in Congress, rather than when might be predicted if groups were perfect agents of their members or other principals. Thus, interest group behavior and influence are a product of strategic dynamics, and MapLight data allow for large-scale examination of these dynamics.

Finally, researchers can also leverage the MapLight data to produce summary measures related to interest group and lawmaker behavior and congressional institutions. As noted above, Fagan et al. (2019) combine the MapLight data with information about individual bills to develop new measures of both the alignment of a group with one of the two major congressional parties as well as the agenda (i.e., issue area) diversity of the bills on which a group took a position. Crosson et al. (2019) apply IRT scaling methods to estimate ideal points for over 2600 interest groups and 950 members of Congress. The scores point toward a "polarized pluralism" among interest groups in Washington, in which groups have arrayed themselves in two distinct leftward and rightward modes—much like Congress. These advances notwithstanding, potential applications of the MapLight data transcend the study of interest group politics. Using the timing-specific version of the data made available by Holyoke (2019), for example, Crosson et al. (2019) estimate proposal and status quo locations for congressional bills. Their preliminary findings show that while the legislative process does tend to weed out extreme proposals, it does not appear that partisan agenda-setting powers are the direct mechanism of this effect. Thus, MapLight data can be marshalled to reexamine longstanding theories of American politics, both those specifically related to interest groups and to other questions about the health of American political institutions.

Taken together, the breadth of current research using the MapLight Bill Positions dataset underscores its potential for application to a wide variety of analyses. Given the recency of many of the existing applications, it is likely that there is much yet to learn from examining these data. The remainder of this paper therefore describes the MapLight data in greater detail and discusses potential strengths and limitations with its use. We then explain how to acquire the data and discuss ongoing challenges for applying it broadly to research on interest group politics. We conclude by pointing interested readers toward sources of position-taking data from other governmental contexts.

MapLight bill positions data contents and coverage

Here, we describe the contents and coverage of the data in MapLight's Bill Positions API, with regard to both groups and bills. Each entry in this dataset is at the Congress-Bill Number-Organization level. The dataset contains the following variables:



- *url*: The web address for the page on MapLight’s website where the bill can be viewed individually.
- *session*: the Congress in which this bill was introduced.
- *prefix*: the bill’s prefix, indicating its chamber of origin (e.g., it uses H for normal House bills and S for Senate bills).
- *number*: the bill’s bill number.
- *measure*: the bill’s full designation.
- *topic*: the bill’s name, as specified in bill text.
- *MLid*: a unique¹ numerical identifier for the organization taking the position.
- *orgname*: the name of the organization taking the position.
- *disposition*: the position taken by the organization on the bill.
- *citation*: a citation for a web url to the source by which MapLight recorded the position. Note that this citation includes the dates both when the cited source was published as well as when MapLight accessed that source to record the position(s) it contained.
- *grouptype*: the catcode, per the interest group taxonomy created by the Center for Responsive Politics (CRP),² assigned by MapLight to represent the organization’s interest with respect to a particular bill.

The bill-identifying fields (*session*, *prefix*, *number*) allow the MapLight data to be merged with datasets more commonly used in political science, such as the Congressional Bills Project (<http://www.congressionalbills.org>, accessed 21 October 2019). Moreover, a combination of the *session* and *catcode* variables potentially allow a researcher to integrate the MapLight data with catcode-level data on campaign contributions or other expenditures (e.g., Lorenz 2020). Thus, the MapLight data can in principle be used to integrate information on interest groups’ bill positions into studies of many aspects of legislative and electoral politics.

The MapLight data feature a broad set of interest groups in the Congresses MapLight has researched. Across the 109th–115th Congresses (2005–2018), MapLight has (as of May 2019) recorded 130,707 positions taken by 16,555 unique organizations on 9979 bills. Table 1 depicts the distribution of positions by the CRP Sector (an aggregation of the *grouptype* variable described above) and the distribution of the organizations themselves. This table demonstrates that the organizations taking these positions represent a broad swath of different types of industries. The modal group is an ideological or single-issue group, likely reflecting the many small nonprofit organizations who, barred from traditional lobbying, take public positions on legislation as a signal of activity to their members. On the other hand, the data do not as frequently capture groups that might lobby on segments of much larger bills. Perhaps the canonical example of such groups

¹ In our analysis, some groups with identical names appear to be associated with more than MLid. This may result from idiosyncrasies in naming conventions used by MapLight’s records, and the Center for Responsive Politics (CRP) conventions on which they are built, particularly as they track organizations’ formation, merging, acquisition, and dissolution. In a future release of the data, we plan to resolve these errors.

² <https://www.opensecrets.org/industries/slist.php>, accessed 21 October 2019.



Table 1 Distribution of positions by CRP sector of the organization

	Count of positions	Share of positions	Count of groups	Share of groups
Ideology. Single. Issue	45,171	0.35	4998	0.25
Misc. Business	14,903	0.11	2180	0.11
Health	14,346	0.11	2253	0.11
Other	11,313	0.09	2151	0.11
Agribusiness	7692	0.06	1108	0.06
Labor	6993	0.05	343	0.02
Communic. Electronics	6384	0.05	1182	0.06
Finance. Insur. RealEst	5834	0.04	857	0.04
Energy. Nat. Resource	4740	0.04	936	0.05
Construction	3855	0.03	432	0.02
Transportation	3054	0.02	556	0.03
Lawyers & Lobbyists	643	0.00	114	0.01
Defense	357	0.00	56	0.00
Campaign finance related	58	0.00	18	0.00
NA/Unknown	5364	0.04	2749	0.14

The total count of groups represented above sums to more than the total number of unique groups in the dataset, because MapLight codes each grouptype at the group-bill level, attempting to best capture the interest that group represents with regard to that particular bill. As such, some groups are categorized in several groups and are thus represented multiple times in the aggregation above

might be the defense industry, who lobby intensely for increased defense spending in appropriations but only infrequently take positions on other legislation.

With regard to the valence of groups' position-taking activities, MapLight data capture considerably more support of than opposition to legislation. Indeed, roughly 74% (96,794) of the positions are instances of an organization supporting a bill, while the remainder (33,913) are expressions of opposition. While it is not possible with existing data sources to know to what extent this is representative of lobbying as a whole, it is worth noting that the 74% support decisions by groups closely reflects roll-call patterns on the bills in the dataset: roughly 75% (310,373 out of 409,188) of votes cast on bills found in the MapLight data were "yay" votes in Congress. These similarities between roll-call voting and interest group position-taking obtain throughout the policymaking process, as captured in Fig. 1. Moreover, this roughly 3-to-1 ratio is consistent with McKay's (2012) finding regarding the (im-)balance of support and opposition generally necessary among the lobbying community for the adoption of a policy proposal. Further still, as Baumgartner et al. (2009) underscore, groups primarily interested preserving the status quo—that is, opposing policy change—frequently husband their resources and abstain from advocacy until the bill constitutes a viable threat to the status quo. For this reason, groups in opposition to legislation are unlikely to take positions on bills that die in committee. These positions constitute a sizeable portion of the MapLight data (again, as shown in Fig. 1). Thus, while the



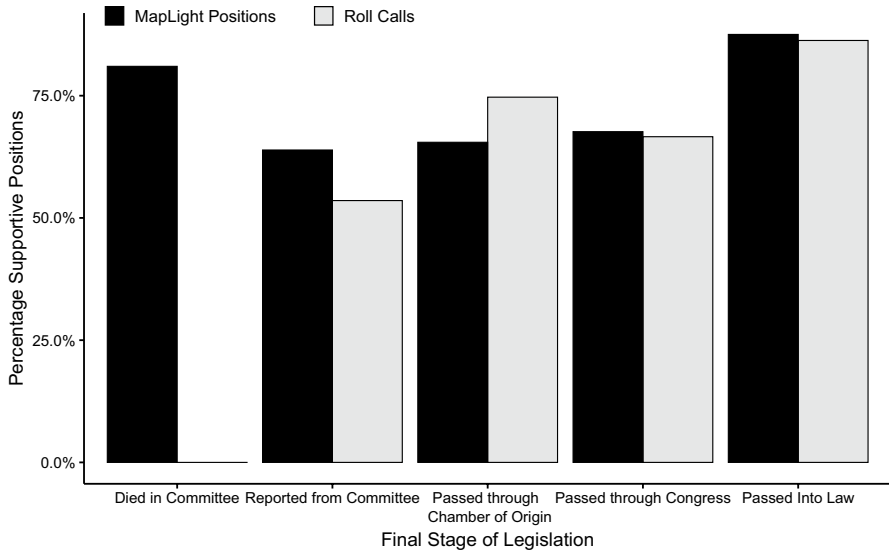
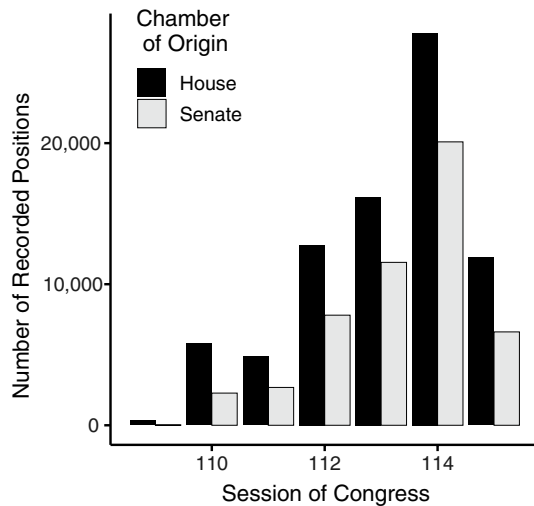


Fig. 1 “Support” Positions Among Interest Groups and Members of Congress, by Bill Stage. *Note* Because the Congressional Bills Project extends only through the 114th Congress, this graph does not include the 115th Congress data now available through MapLight

Fig. 2 Positions recorded by MapLight on bills in the 109th–115th Congresses as of May 2019 date



balance between expressed support and opposition is uneven, it is perhaps also unsurprising and still representative of position-taking as a whole.

In addition to covering an impressive variety of groups, MapLight’s coverage of legislation is also quite extensive in several contexts relevant to applied research. The first such context simply relates to the recency of the data: MapLight appears to have improved their data collection process since their founding. Figure 2 depicts



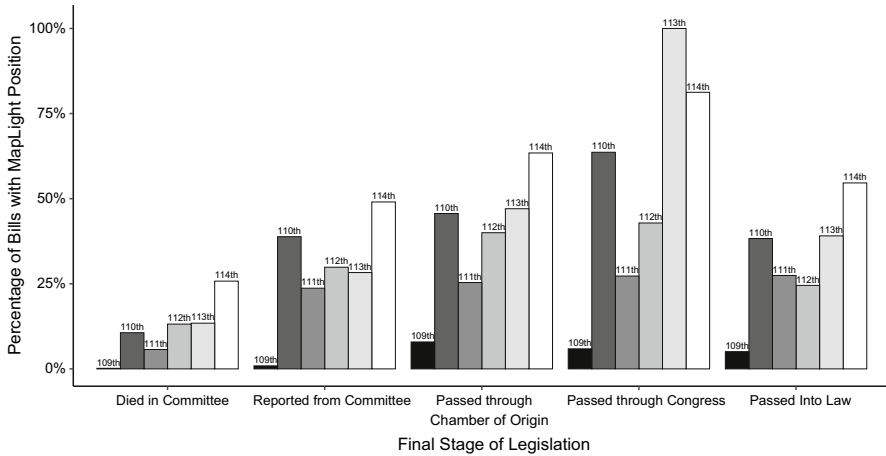


Fig. 3 MapLight data coverage, by bill outcome. *Note* Because the Congressional Bills Project extends only through the 114th Congress, this graph does not include the 115th Congress data now available through MapLight

the distribution of positions by session of Congress, as well as by chamber of origin associated with the bills for each position taken. Positions are generally more numerous in recent Congresses, regardless of the chamber of origin. Additionally, MapLight has found more positions on House bills (79,652 positions recorded across the time period) than Senate bills (51,055). However, if one assumes that House and Senate bills are equally likely to generate positions, these position-taking figures track reasonably well with the number of bills introduced in each chamber over this period (46,905 and 26,011, respectively).

The second context within which MapLight data coverage is extensive is on legislation that has moved forward in the legislative process. As Fig. 3 summarizes, for all Congresses in the dataset, bills that die in committee appear much less likely to attract position-taking, MapLight research, or both, compared to those bills that progress past the committee stage. In fact, in the most extensive Congress (the 114th), MapLight records positions on only 25% of bills that die in committee. Meanwhile, for bills that gain more traction, MapLight records positions for as many as 80–100% of bills, depending on the year and level of progress. Thus, for scholars interested in later-stage bills, MapLight's data provide impressive coverage. It should be noted, though, that this relationship does not appear to be monotonically positive, as MapLight's coverage dips for bills that ultimately become law. It is unclear, however, as to the percentage of such bills that enjoy unanimous support in Congress and are thereby unlikely to attract interest group position-taking.

Finally, MapLight positions are well distributed across issue areas. Indeed, as depicted in Table 2, there are well over a thousand positions (and often several thousand) taken on bills in almost every Comparative Agendas Project Major Topic Code. Furthermore, the distribution of this position-taking activity captured in the MapLight data is quite similar to the distribution of lobbying activity according to bills mentioned in Lobbying Disclosure Act (LDA) activity reports. Despite the



Table 2 Distribution of positions by the CAP Major Topic Code of the bill

Major topic code	Count of positions in MapLight data	Share of positions in MapLight data	Share of occurrences in LDA data
Macroeconomics	4651	0.04	0.03
Civil rights	5590	0.05	0.03
Health	16,834	0.15	0.17
Agriculture	4356	0.04	0.03
Labor	6013	0.06	0.03
Education	4608	0.04	0.03
Environment	5910	0.05	0.05
Energy	5942	0.05	0.08
Immigration	3273	0.03	0.01
Transportation	3461	0.03	0.04
Law & Crime	5555	0.05	0.04
Social welfare	2961	0.03	0.02
Housing	1075	0.01	0.01
Domestic commerce	12,461	0.11	0.11
Defense	3189	0.03	0.09
Technology	2756	0.03	0.04
Foreign trade	4314	0.04	0.02
International Affairs	1501	0.01	0.01
Governmental operations	11,393	0.10	0.13
Public lands	3452	0.03	0.05
Culture	28	0.00	0.00

Activity on bills that did not receive a major topic code from the Comparative Agendas Project have been excluded. As a reference, the share of activity listed on lobbying disclosure reports that is directed toward bills in each topic codes is listed in the rightmost column. Organizations mentioning specific bills multiple times on LDA activity reports were counted only once per bill

overall similarity in these two categorical distributions shown in Table 2, Maplight data do tend to over-represent positions taken on labor legislation and under-represent those taken on defense, governmental operations, energy and health, to some degree. Nevertheless, it is worth noting that while the MapLight positions may be well distributed across chambers and issue areas, they are not proportionate across parties: 75.56% of positions recorded by MapLight were taken on bills introduced by members of their chambers' majority party. Fortunately, over the time period for which MapLight data exist, both Republicans and Democrats enjoyed majority control of the House and Senate for at least one Congress.

Overall, the set of bills for which there is MapLight position-taking is rich and expansive across many features of interest to applied research. In spite of the impressive coverage the MapLight data exhibit, there remain at least two important caveats. First, MapLight does not represent a census of all interest group position-taking. This implies that there are an unknown number of cases in which groups



took a position on the bill but MapLight did not record it. To the extent that these “false non-positions” are non-random across group types of interest to a particular descriptive or causal research question, MapLight data may produce less valid inferences. Second, interest groups do not take positions on all bills before Congress. As a result, many bills in Congress are not included in the MapLight data, and the subset that are included are unlikely to be a random sample of bills. Thus, we turn next to a more systematic analysis of the factors that make a bill more likely to generate positions MapLight records.

MapLight bill selection

Perhaps the consideration most likely to affect the usefulness of MapLight data for social science research is sample selection bias. As noted above, MapLight data are not a census of interest group activity. Indeed, the 9979 bills for which MapLight identified public position-taking between the 109th and 115th Congresses constitute less than twenty percent of all bills introduced during this period. It is unlikely that interest groups did not lobby and even that they did not take public positions on other bills. In addition to being a smaller sample than might be ideal, MapLight does not randomly select bills for research. In its current (as of October 2019) documentation for the application programming interface (API) by which its position-taking data can be acquired (see below), MapLight simply states that it does not attempt to research commemorative bills (e.g., those that concern “resolutions for commemorative postage stamps” or “National Life Insurance Awareness Month”. See https://MapLight.org/data_guide/bill-positions-api/, accessed 16 October 2019). However, it is possible that factors for which MapLight does not intentionally select nevertheless influence the likelihood that a given bill is chosen for research. Similarly, there may be factors that affect whether MapLight is able to locate positions for a bill. Either issue could introduce sample selection problems into MapLight’s data collection process, hampering the data’s potential for generating descriptive or causal inferences.

We therefore systematically examine the consequences of these features of MapLight’s research process for the use of MapLight’s data in scholarly research. Specifically, we estimate a logistic regression model of inclusion in MapLight’s position-taking data, as a function of many factors commonly examined in scholarship on Congress. A key advantage of doing so that it allows us to both confirm biases MapLight explicitly cites as guiding their selection (e.g., bills that are non-commemorative), but also factors that may not explicitly guide MapLight’s research process but nonetheless influence the likelihood that they research a bill for interest group positions.

In this analysis, we identify each bill in the MapLight dataset for which MapLight has identified one or more organizations as having taken a position supporting or opposing the bill. Importantly, this measurement captures both MapLight’s decision to research a bill and the result of having found any position-taking. This indicator variable—which equals 1 if MapLight found interest group positions and 0 otherwise—is then joined to the subset of the Congressional Bills Project dataset



covering the years in which MapLight researched bill positions. We regress this variable onto a wide variety of factors potentially associated with MapLight's decision to research positions on a bill. Given the number of terms in the regression, we caution that some coefficients may be attenuated. Thus, the results reported below should be taken as suggestive, with other potential factors possibly masked by the inclusive nature of the empirical model.

Results are reported in Figs. 4 and 5, which depict the logit coefficients and confidence intervals for the main effects and the issue-area fixed effects, respectively. They are consistent with MapLight's most recently stated summary of its bill selection procedures, in that the coefficient on *Important Bill*, which indicates whether a bill is not a commemorative bill,³ is positive, larger than the coefficient for any other dichotomous variable in the model, and statistically significant. However, there are several other factors that also have made it more likely that MapLight researched and found positions on a given bill. Though House bills (H, the excluded category in the model)⁴ account for a larger proportion of bills in the Congressional Bills Project dataset than Senate bills, bills originating in the Senate are actually more likely to be included in the MapLight dataset. Various types of resolutions in either chamber are less likely than normal House and Senate bills to be MapLight bills. Bills introduced in later Congresses, that have large numbers of cosponsors, that are multiply referred, or that attain some level of legislative advancement (or that get passed by Congress and then vetoed) are also more likely to be MapLight bills. Finally, bills introduced by members in the majority party, who are ideologically liberal, who are ideologically extreme, or that have relevant institutional positions (e.g., being a member or chair of the committee to which the bill was referred) are more likely to be MapLight bills. Turning to Fig. 5, it also appears that MapLight finds positions more often among bills in some issue areas than others. Together, these results suggest that MapLight's research process is influenced by a variety of bill- and sponsor-level characteristics.

These clear biases undoubtedly raise important concerns about the extent to which MapLight's research decisions may affect inferences made using MapLight data. On the whole, however, we believe that if analysts are careful, these effects can be minimized. We therefore consider how researchers interested in making descriptive and causal inferences about interest group activity may consider using MapLight data.

³ Volden and Wiseman (2014) differentiate between commemorative bills, substantive bills, and "substantive and significant" bills. The latter category is measured as the bill having received a write-up in an end-of-year Congressional Quarterly Almanac or, for more recent Congresses, the Congressional Quarterly Weekly/CQ Magazine during the Congress in which it was introduced. While we do not require this distinction to demonstrate MapLight's preference for researching non-commemorative bills, it may prove useful for scholars working with MapLight data to leverage this more granular measure of legislative significance.

⁴ Readers with knowledge of congressional bill formatting may be aware that this is somewhat atypical of how Congress itself abbreviates bill types. In the MapLight data, and thus by implication in Fig. 4, "H" is used for House Bills (typically abbreviated "H.R."), while "HR" is used for House Resolutions (typically abbreviated "H.Res.").



Maplight Selection Main Effects

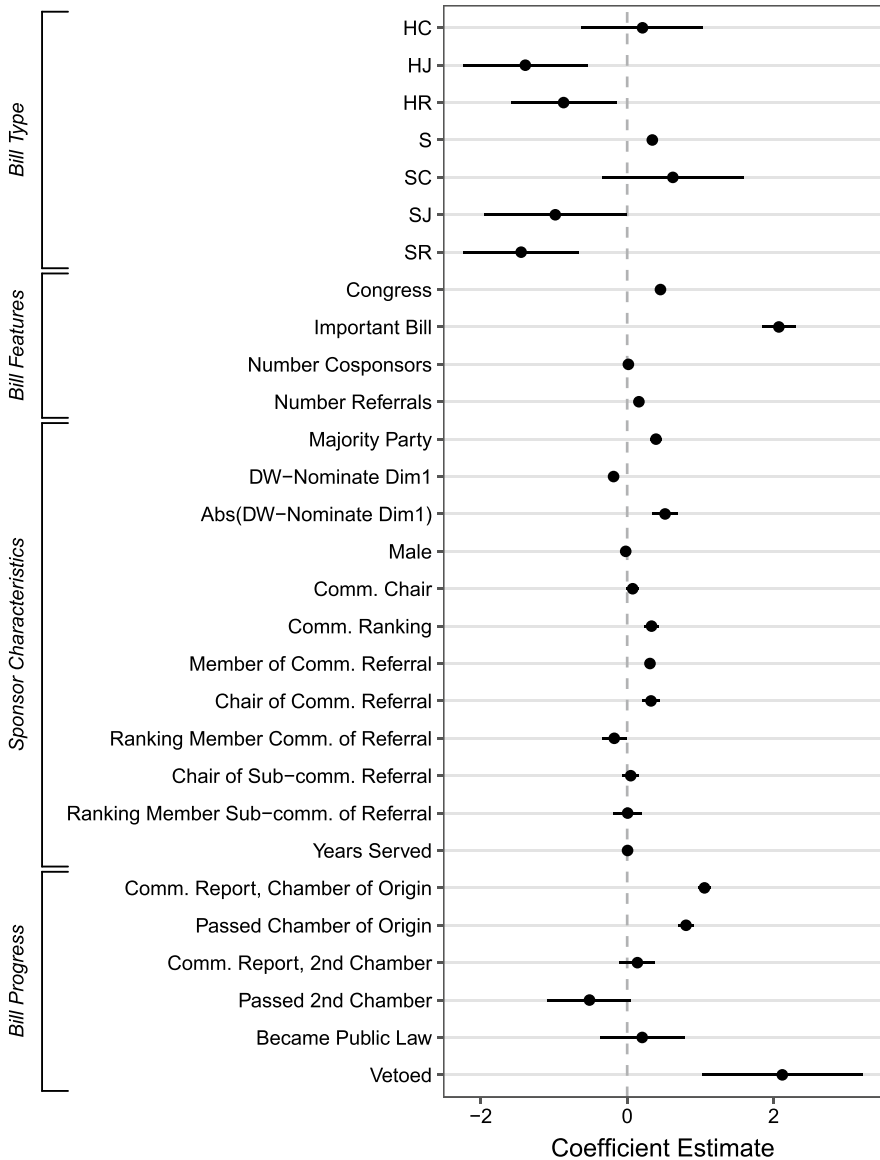


Fig. 4 Coefficient plot for logit model of whether a bill is included in the MapLight bill positions data. Note “H” bills are excluded category in Bill Type

First, with regard to making descriptive inferences about the population of all bills from MapLight bills, we believe researchers should focus their inquiries on contexts in which a critical component is not just interest group *activity* (which may alternatively be captured by, e.g., lobbying disclosures), but organizations’



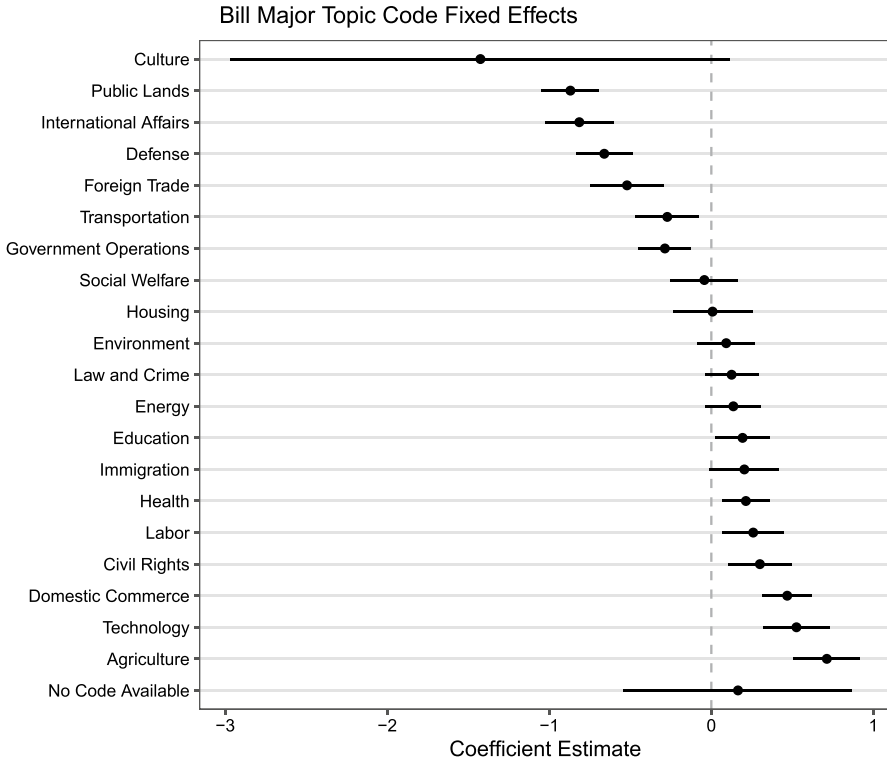


Fig. 5 Coefficient plot for logit model of whether a bill is included in the MapLight bill positions data, continued. *Note* Macroeconomics is the excluded category of issue areas

“support/opposition” *positions*. Even in that case, analysts should clarify that they are drawing descriptive inferences about a particular set of bills, and note the sampling biases in the MapLight dataset relative to the known properties of the congressional bill population. Because MapLight bills are a subset of all bills, and because we have comprehensive information about the total population of bills, analysts may apply post-stratification weights based on the full population of bills to make more generalizable descriptive inferences.

Causal inferences are also potentially affected by systematic influences on the selection of MapLight bills. In particular, inferences about the population of bills that are made using the MapLight data are liable to various forms of sample selection bias. These may be addressed by using statistical models that explicitly account for factors that influence the sampling process (i.e., Heckman and related models). Even absent the use of such models, however, it is worth considering what the effects of MapLight’s sampling strategy could have within particular applications. Conceptually, the effect of selecting on a variable is to reduce the probability that observations with values of that variable near one end of its range appear in the sample (in the extreme case, by censoring observations above or



below a particular value of that variable). This is of particular concern when one of MapLight's selection criteria (both stated and unstated) are strongly related with an analyst's independent or dependent variable of interest. The precise effect of such selection problems depends upon the role of the variable in the causal model (King et al. 1994). If the independent variable, then the effect of selection bias is, in expectation, negligible. If it is the dependent variable, however, the effect is to attenuate the estimated effect size, as the relative absence of higher or lower values "flattens" the slope of estimated the regression line. Thus, when a variable affecting MapLight's sampling strategy is also the dependent variable in an analysis, the estimated effect size is smaller than it would be in the larger population of bills. We expect this to be particularly relevant for studies attempting to use MapLight data to assess the effects of interest group advocacy on patterns of cosponsorship, legislator ideology (particularly as revealed through roll-call votes), and bill advancement.

Thus, MapLight's research strategy can affect inferences drawn using its bill positions data. Though analysts can account for these effects as outlined above, it is also possible that the "biases" of MapLight's sampling strategy are useful for some applications. Indeed, MapLight's sampling strategy filters the population of bills for those with a set of attributes—e.g., likelihood of containing policy substance and actual potential to advance in the legislative process—that may be useful for a particular analysis. For example, it may be desirable to distinguish between bills offered purely for messaging purposes (i.e., with their sponsors neither anticipating nor pursuing actual legislative advancement) from those offered with the intent and actual possibility that they may pass into law. If a particular study wishes to focus on the latter bill type, or simply distinguish between them, then MapLight bills may proxy for the theoretical population of interest. We caution, though, that the validity of doing so will depend on the particular application. Authors should therefore be expected to explicitly state their group- or bill population of interest and explain why MapLight bills allow them to make generalizations accordingly.

Other considerations

In addition to the more traditional sampling issues outlined above, there are several aspects of MapLight's research process that may affect the utility of its bill positions data for scholarly research.

Many explicit positions on many specific bills. Though we demonstrate above the wide range of potential uses for MapLight's data, it is worth noting how these data compare to alternative large-scale data sources for identifying interest group activity in Congress. In particular, that MapLight captures support and opposition rather than merely interest and activity, renders MapLight data a useful alternative to data drawn from Lobby Disclosure Act (LDA) filings. LDA filings have improved considerably since their original introduction, particularly with regard to specificity about bills of interest. Nevertheless, reporting of positions remains rare. Moreover, MapLight data are unique among large-scale interest group activity datasets in capturing many instances not only of support but also of opposition. Potential alternative data sources, such as legislators'



press releases or Dear Colleague letters, also infrequently report opposition to bills, since they are usually produced by a bill's sponsor or that sponsor's collaborators.

Organizational population. Many organizations attempt to influence Congress but neither report lobbying nor give campaign contributions from affiliated PACs. Because of this, they cannot appear in datasets that rely solely on LDA reports or PAC expenditures. While some types of important advocacy organizations—e.g., think tanks and 501(c)(3) nonprofits—are legally barred from lobbying, many other types of organizations (e.g., individual firms) do not hire lobbyists directly and instead rely on organizations like trade associations to lobby on their behalf. In many cases, these organizations' primary function is not necessarily to lobby, but they nonetheless take positions on particular legislative proposals and thus potentially influence the politics surrounding those proposals. This inclusivity is generally an advantage of using MapLight data, but it implies that analysts considering the MapLight data should examine whether their conceptual definition of an "interest group" aligns with MapLight's inclusion of organizations taking positions regardless of their primary organizational form and functions.

MapLight's research capacity has improved over time. As MapLight has matured as an organization, it has more frequently found interest group positions. As demonstrated in Fig. 2 and implied by the positive coefficient on Congress in Fig. 4 above, while MapLight recorded only a handful of positions in the 109th Congress, and roughly 10,000 positions in the 110th, for the 114th it recorded over 50,000 (roughly 40% of all positions MapLight has recorded and released through its API). This suggests that bills are more likely to be researched successfully in more recent Congresses, and thus results using the MapLight data are more reflective of recent Congresses than earlier Congresses. Analysts might consider weighting observations to account for this, depending on the application. However, given MapLight's coverage across Congresses, such weightings are likely to be more successful in producing representative estimates in Congresses from the 110th onward.

Timing of positions. Part of the MapLight data's "citation" field (see above) records the publication date of the source MapLight used to identify the position of a group on a bill. It is not obvious from these citations or the documentation how MapLight accounts for potential changes in groups' positions. According to direct correspondence with MapLight's researchers, however, MapLight's Bill Positions API contains only the *most recent* published position taken by an organization on a bill (the latter identified by its bill number, not its title). Scholars may assume, then, as Crosson et al. (2019) do, that a position taken on an earlier version of the bill applies to subsequent legislation with the same bill number within a Congress. In many cases, this assumption is likely inconsequential,⁵ but it would be less so if a bill received a substitute that substantially altered the policy content of the bill.⁶

⁵ Assuming a unidimensional spatial model of group preferences and that a subsequent version of a bill did not change which side of the status quo that bill fell upon, the organizations for whom such an assumption would not hold would be those who were close to indifferent between the original bill and the status quo it amended.

⁶ For example, before it became the Patient Protection and Affordable Care Act (ACA), 111 H.R. 3590 passed the House as the "Service Members Home Ownership Tax Act of 2009", which amended the tax code to facilitate home ownership by military families.



Crosswalking with other datasets. Though MapLight assigns numerical identifiers to organizations for which it finds positions, there does not yet exist a comprehensive crosswalk from MapLight identifiers to other organizational identifiers such as those provided by the Center for Responsive Politics, LDA reports, or PAC registrations. Crosson et al. (2019) built a crosswalk for the subset of 2646 organizations in the MapLight data that took enough positions for their ideal point to be estimated. They found affiliated PACs for approximately 30% of this organization subset and lobbying disclosures for approximately 57%. To do so, they matched organization names using text analysis techniques of varying sophistication (see, e.g., fn 15 and 23), but even so over ten thousand MapLight organizations remain unmatched to other sources of data about organizational advocacy activities, as of this writing. Fortunately, because MapLight data are specific to particular congressional bills, datasets at the bill- and sponsor-level (e.g., the Congressional Bills Project, Comparative Agendas Project, and Center for Effective Lawmaking) can be readily joined to MapLight data.

Obtaining the MapLight bill positions data

Using the framework given above should give researchers a sense of whether, or with what caveats, the MapLight data may permit them to credibly answer a given research question. This section describes two methods by which the MapLight Bill Positions dataset may be acquired.

First, and our recommendation, is for researchers to download the data directly from MapLight's API. Doing so has the principal advantage of allowing the researcher to obtain the most up-to-date position-taking data, from the 109th to the 115th Congresses (and, we presume, future Congresses). However, it also requires that the researcher process the raw xml or json files that the API generates. Documentation for the API can be found at https://MapLight.org/data_guide/bill-positions-api-documentation/ (accessed 16 October 2019). In particular, we have made use of the API's "Bill List" method, which can return, for each bill in a given Congress, the list of organizations taking positions for or against that bill.

Second, we offer a version of the MapLight data that we have processed, and that was used in the generation of the tables, figures, and regression models presented in this manuscript. This version, and documentation for it, can be downloaded from the replication files attached to this manuscript. This version of the MapLight data was downloaded and processed in May 2019. It contains over 130,000 rows, each representing one organization's position on one bill during one of the 109th–115th Congresses. It contains the variables described in the Data Contents section, above.

Conclusion

In this article, we have sought to introduce the MapLight Congressional Bill Positions data, as made available through its API, and provide a guide to how features of the data and how they are collected may impact scholarly research that uses them. Much of this discussion has focused on what might be viewed as problems with



the MapLight data. Of particular note is MapLight's research process, which has improved over time but nonetheless selects bills for research in a non-random fashion; this results in the MapLight-researched bills being unrepresentative of the full population of congressional bills, in ways that may bear upon common foci in the study of interest groups and Congress. Moreover, at present there is much work needed to be done to connect MapLight groups to the same groups in lobby disclosure and campaign finance data. Given our focus on these issues, it may appear that we are implicitly warning against the use of the MapLight data for scholarly research. This is not our intent. Indeed, we have found the MapLight data invaluable in our own research. Rather, we have sought here to guide analysts toward a set of important factors when considering a use of the MapLight data.

There are alternative sources of group position-taking data covering both different types of policymaking activity as well as different geographical contexts. Osgood (2017) collected and analyzed an original dataset on industries' support and opposition specific to trade liberalization proposals. Similar support/opposition position-taking also occurs on other types of policymaking decisions, including amicus briefs before the Supreme Court (Box-Steffensmeier et al. 2013), as well as bureaucratic agency rulemaking (Dwidar 2019). There are also important efforts to collect and expand position-taking data beyond the U.S. federal government. American sub-national governments provide an especially promising venue for observing interest group influence and institutions' impact on that influence (Anzia 2019), and relatively strong lobbying disclosure requirements in some states (such as Wisconsin, Colorado, Iowa, and Nebraska) provide excellent opportunities to observe groups' support and opposition of legislation in those states (Thieme 2019; Garlick 2016). For their part, MapLight also collects bill positions data for the California State Legislature.⁷ Finally, the GovLis project (particularly, Junk 2019) has collected advocate position-taking for issues across several European countries using surveys. These projects, as well as our own ongoing efforts to supplement the MapLight database, will only expand the available position-taking data that analysts may use.

Like any scholarly tool, the MapLight data are not perfect for all purposes. They do, however, hold much promise. The ability to track the revealed preferences of groups with respect to specific bills has wide potential applicability to studies of interest group behavior and influence as well as of Congress more broadly. Though other types of position-taking are also producing useful data, the MapLight data are, to the best of our knowledge, at present the only publicly available database of interest groups' support and opposition to a wide range of congressional bills as expressed by the groups themselves. Scholars are just beginning to tap into the richness of these data. We hope this guide fosters that work.

⁷ While we have not used the California data and thus cannot speak to its usefulness for applied research, we note that California's lobbying disclosure laws appear to require even less specificity about the issues and bills lobbied than does the federal LDA. Thus, MapLight's California dataset may have fewer available substitutes than does the federal dataset that has been our focus here. In light of this possibility, we look forward to future assessments of the California dataset's validity and reliability.



Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

- Anzia, Sarah F. 2019. Looking for influence in all the wrong places: How studying subnational policy can revive research on interest groups. *The Journal of Politics* 81(1): 343–351.
- Baumgartner, Frank R., and Beth L. Leech. 2001. Interest niches and policy bandwagons: Patterns of interest group involvement in national politics. *The Journal of Politics* 63(4): 1191–1213.
- Baumgartner, Frank R., Jeffrey M. Berry, Marie Hojnacki, Beth L. Leech, and David C. Kimball. 2009. *Lobbying and policy change: Who wins, who loses, and why*. Chicago: University of Chicago Press.
- Bertrand, Marianne, Matilde Bombardini, and Francesco Trebbi. 2011. *Is it whom you know or what you know?* Technical report National Bureau of Economic Research: An empirical assessment of the lobbying process.
- Box-Steffensmeier, Janet M., Dino P. Christenson, and Matthew P. Hitt. 2013. Quality over quantity: Amici influence and judicial decision making. *American Political Science Review* 107(3): 446–460.
- Broz, J. Lawrence. 2016. “China, Currency Misalignments, and Industry Demands for Trade Protection in the United States.” *working paper*. Available at SSRN: <https://ssrn.com/abstract=2490648>.
- Crosson, Jesse M., Alexander C. Furnas and Geoffrey M. Lorenz. 2019. “Estimating proposal and status quo locations for legislation using cosponsorships, roll-call votes, and interest group bill positions. Presented at the annual meeting of the Society for Political Methodology (POLMETH), Boston, MA, July 18–20, 2019.
- Crosson, Jesse M., Alexander C. Furnas and Geoffrey M. Lorenz. 2019. “Polarized pluralism: Organizational preferences and biases in the American pressure system”. *American Political Science Review*. accepted.
- Dwidar, Maraam A. 2019. “Diverse lobbying coalitions and influence in notice-and-comment rulemaking”. *working paper*.
- Fagan, E. J., Zachary A. McGee and Herschel F. Thomas. 2019. “The power of the party: Conflict expansion and the agenda diversity of interest groups.” *Political Research Quarterly*. accepted.
- Furnas, Alexander C., Michael T. Heaney, and Timothy M. LaPira. 2019. The partisan ties of lobbying firms. *Research & Politics* 6(3): 1–9.
- Galantucci, Robert A. 2015. The repercussions of realignment: United States-China interdependence and exchange rate politics. *International Studies Quarterly* 59(3): 423–435.
- Garlick, Alexander Russell. 2016. “Interest groups, lobbying and polarization in the United States”, PhD thesis, University of Pennsylvania. 2298. <https://repository.upenn.edu/edissertations/2298>
- Grossmann, Matt, and Kurt Pyle. 2013. Lobbying and congressional bill advancement. *Interest Groups & Advocacy* 2(1): 91–111.
- Holyoke, Thomas T. 2019. Strategic lobbying to support or oppose legislation in the U.S. congress. *Journal of Legislative Studies* 25(4): 533–552.
- Junk, Wiebke Marie. 2019. When diversity works: The effects of coalition composition on the success of lobbying coalitions. *American Journal of Political Science* 63(3): 660–674.
- Kim, In Song. 2017. Political cleavages within industry: Firm-level lobbying for trade liberalization. *American Political Science Review* 111(1): 1–20.
- Kim, In Song. 2018. “LobbyView: Firm-level lobbying & congressional bills database.” Working Paper available from <http://web.mit.edu/insong/www/pdf/lobbyview.pdf>.
- King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing social inquiry: Scientific inference in qualitative research*. Princeton: Princeton University Press.
- LaPira, Timothy M., and Herschel F. Thomas. 2017. *Revolving door lobbying: Public service, private influence, and the unequal representation of interests*. Lawrence: University Press of Kansas.
- Lorenz, Geoffrey Miles. 2020. Prioritized interests: Diverse lobbying coalitions and congressional committee agenda setting. *The Journal of Politics* 82(1): 225–240.



- McKay, Amy. 2012. Negative lobbying and policy outcomes. *American Politics Research* 40(1): 116–146.
- Moore, Ryan T., Eleanor Neff Powell, and Andrew Reeves. 2013. Driving support: Workers, PACs, and congressional support of the auto industry. *Business and Politics* 15(2): 137–162.
- Osgood, Iain. 2017. The breakdown of industrial opposition to trade: Firms, product variety, and reciprocal liberalization. *World Politics* 69(1): 184–231.
- Thieme, Sebastian. 2019. “Moderation or Strategy? Political Giving by Corporations and Trade Groups.” *Journal of Politics*. forthcoming.
- Vidal, Jordi Blanes I., Mirko Draca, and Christian Fons-Rosen. 2012. Revolving door lobbyists. *The American Economic Review* 102(7): 3731–3748.
- Volden, Craig, and Alan E. Wiseman. 2014. *Legislative effectiveness in the United States congress: The lawmakers*. Cambridge: Cambridge University Press.
- You, Hye Young. 2017. Ex post lobbying. *The Journal of Politics* 79(4): 1162–1176.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

