Analysis of Cross-border Data Trade Restrictions using Mixture-based Clustering

Keman Huang, Samuel Chinnery, Stuart Madnick

Working Paper CISL# 2019-08

March 2019

Cybersecurity Interdisciplinary Systems Laboratory (CISL)
Sloan School of Management, Room E62-422
Massachusetts Institute of Technology
Cambridge, MA 02142

Analysis of Cross-border Data Trade Restrictions using Mixture-based Clustering

Completed Research

Keman Huang
Cybersecurity at MIT Sloan, Sloan School of Management, MIT
keman@mit.edu

Samuel Chinnery
Electrical Eng. & Computer Sci., MIT
chinnery@mit.edu

Stuart Madnick
Cybersecurity at MIT Sloan, Sloan School of Management & MIT School of Engineering
smadnick@mit.edu

Abstract

Cross-border data flows and interoperability have created an increasingly digital global trade environment. Concerns for cybersecurity, privacy, and digital sovereignty have prompted the widespread implementation of trade policies regulating cross-border data flows. At present, there is a deficit of information characterizing such policies and analyzing their role in the global economy. We address this deficit through a quantitative analysis of trade policies in services that seeks both to categorize data-related trade policies and to evaluate the extent to which they are different from other restrictive measures. We propose a mixture-based clustering pipeline to group trade restrictiveness data and a method for quantifying the difference between the cross-border data flow regulations and other traditional regulations. Our analysis reveals a significant localization effect among data flow restrictions and that, while highly restrictive data flow policies generally do not overlap with other policies, there is a significant similarity in moderate to liberal policies.

Keywords

Digital Trade, Trade in Services, Cross-border Data Flows, Trade Restriction Policy Analysis.

Introduction

World exports of commercial services exceeded $5 trillion USD in 2017 (Koopman and Maurer 2018). Trade in services has been growing at an unprecedented rate, spurred by new developments in science and technology and increases in demand for data and intellectual property (IP). Of all commercial services, trade in IP grew the most in 2017, bringing concerns for data privacy and data security to the forefront. With the advent of the EU General Data Protection Regulation (GDPR) and other data-related restrictions, there is a growing necessity to develop systematic frameworks to analyze these restrictions and the effects they have on global economies (Madnick et al. 2019).

Concerns for cybersecurity, data privacy and digital sovereignty have motivated many countries to implement diverse trade policies that have reshaped the digital trade environment. In the last decade, there has been a trend toward data localization, with many governments and firms seeking to privatize access to the Internet and regulate its use. Several world powers, including Australia, China, and the European Union, have implemented measures to keep data within their borders, with many nations requiring sensitive data to be stored domestically (Chander and Le 2014).

As we will see, these restrictions are often anomalous in the scheme of global trade restrictions, with many countries that are not generally restrictive imposing strict cross-border data flow regulations. Given that international trade is entering the digital trade age, cross-border data flows and interoperability are the cornerstones for digital trade. It is critical to understand the difference between cross-border data flow trade policies and general policies regulating trade in services. This paper aims to quantify these differences between cross-border data flow regulations and trade in service regulations as a whole and to provide a framework for categorizing those policies.

Specifically, we seek to see the extent to which countries are implementing cross-border data flow trade policies that are different from general service trade regulations and to describe the underlying patterns. To that end, we use the OECD (Organization for Economic Co-operation and Development) Service Trade Restrictiveness Index (STRI) database, which will be described in further detail. We propose a mixture-based clustering pipeline to classify countries into different groups based on their trade policies in cross-border data flows and trade in services as a whole. Through this, we will be able to quantify the differences in trade restrictions among different groups of countries. This allows us to further discuss global cyber norm implementation related to cross-border data flow trade policies in the digital environment.

In the following sections, we will summarize the state of the art in data localization policies analysis and then discuss the OECD STRI database, which is used to quantify the trade restrictiveness. The methodology to identify different groups of trade policies is presented. Further analysis is used to reveal the similarity and differences among these groups. We summarize this study with discussion from the analysis.

Data Localization Restriction Policies

Recent changes in international data flow policies are reflective of a broader trend toward "data localization," which has been widely opposed in the literature. Chander and Le (2014) describe this trend in a number of leading economies, arguing that efforts to regulate data flow can slow the progress of major advancements in the field and "erode the ability of consumers and businesses to benefit from access to both knowledge and international markets." They make a distinction between data protection and "data protectionism" and claim that, while data protection is necessary, prohibiting personal data transfer entirely is unjust. Selby (2017) argues further that countries like Russia and China have historically used data localization "to reduce their comparative disadvantage in Internet data hosting" and to reduce the comparative advantage of the US in internet signals intelligence. These efforts have been actively opposed by the US, an economy that brandishes a digital trade surplus in excess of $100 billion.

Cohen et al. (2017) echoes this sentiment, focusing on the distinction between data protection and data localization. Cohen notes that "some commentators speculate that data localization efforts may be less about data protection than trade protection," with many countries using data localization to counteract the rise of multinational internet conglomerates. Others point to the recent rise of cloud computing and its role in decreasing the cost of making computations (Ryan et al. 2013). These authors argue that efforts to use data localization to increase operational security may actually be more harmful than good, with localization efforts in Europe failing to promote government accountability with data protection. In these instances, the governments have "emerged as the principal threat to data privacy."

Unlike these studies focusing on discussion about the negative impact of data localization policies, we seek to understand and quantify differences between trade policy in services as a whole and restrictions in cross-border data flows.

Data Overview: Quantification of Trade Restrictiveness

OECD STRI

Quantifying policy restrictions has been a long-standing concern of economic and policy researchers. As early as 2012, the Trade and International Integration unit of the World Bank's Development Research Group began a project to construct a database of restrictiveness indices. The World Bank's database was based on surveys of developing countries, as well as existing legal data from developed countries. Each policy measure received a score between 0 and 100, with 100 indicating a most restrictive policy measure and 0 indicating virtually no restriction (Borchert et al. 2012). The Organization for Economic Co-operation and Development (OECD) launched a similar project in 2014 aimed at providing a more objective overview of service trade restrictions. The OECD Service Trade Restrictiveness Index database (STRI) examines 44 global economies (36 OECD member countries and 8 other leading economies) and offers an unprecedented depth of information, covering nearly 400 different policy measures across 22 sectors.

Data from the OECD STRI database was gathered through a combination of observational and empirical evidence. Trade experts reviewed regulations for over 40 different countries, dividing laws by sector and categorizing them by policy area. Each observation in the database was verified through peer review, with every country's values reviewed by local policy officials. Each datapoint includes a numerical value and information regarding the source for that value. This allows for more in-depth research into the restrictions if necessary: the source fields provide links to actual trade laws that can offer additional insight.

Each numerical value is derived from a complex scoring methodology. At a high level, the values are the weighted average of a number of binary responses, where each binary response asks a specific policy question. For instance, most sectors in the STRI place limits on foreign equity. This restriction highlights an issue with using binary responses: the limitations on foreign equity can take on any number of values. To solve this problem, OECD researchers created a number of responses for this restriction, with each response indicating a different level of equity limitations. Many other policy measures use scoring of this form, with more "yes" answers corresponding to a higher level of restriction.

The variability of trade restrictions, even within one sector, poses a complex problem for quantification: if each numerical STRI value is to be a weighted average, how should the weights be determined? The OECD project used expert judgment to determine the weights for individual policy measures. In this scheme, policy experts review all policy restrictions and are surveyed to gauge the impact each one has on trade. The judgement process could take several forms: each expert could be allotted "points" they can use to indicate which restrictions are more important. Alternatively, every expert could be asked to rank all the policy measures from most to least significant. In this scheme, care must be taken to ensure many experts agree to ensure objectivity. The latter process was chosen for the final STRI due to its sample independence: weights assigned to each policy measure do not depend on the values of the STRI themselves or on trade balance (Nordås 2009).

The OECD database was chosen for this project by virtue of its depth (more policy measures means easier analysis of a specific sector), objectivity (the data was collected through independent analysis of trade laws), and currency (the database has been updated as recently as January 2019 to include policy data from 2018).

The STRI database covers all 36 OECD member countries, as well as Brazil, China, Colombia, Costa Rica, India, Indonesia, Malaysia, Russia, and South Africa. Data is grouped by sector, and a list of policy measures is associated with each sector to quantify each country's restrictiveness. Each country-sector pair is assigned five values between 0 and 1, with 1 indicating near-complete restriction and 0 indicating near-openness. The five values correspond to the five policy areas: restrictions on market entry, restrictions to the movement of people, barriers to competition and public ownership, regulatory transparency, and other discriminatory measures. Table 1 shows the most and least restrictive economies and sectors in the STRI.

The overall mean STRI is 0.264. When grouped by country, the median STRI is 0.252 and the standard deviation is 0.076. When grouped by sector, the median STRI is 0.238 and the standard deviation is 0.059. In general, there is more variability between countries than between sectors—the range of mean STRI values per country is greater than the range of mean STRI values per sector. The distribution of sector STRI groups is skewed right to a higher degree than the distribution of country groups.

Country	**Mean STRI**	**Sector**	**Mean STRI**
India	0.4877	Air transport	0.4233
Indonesia	0.4603	Legal	0.3840
Russia	0.4467	Accounting	0.3262
China	0.4462	Broadcasting	0.3031
Iceland	0.3849	Rail freight transport	0.2979
...	...	...	...
Ireland	0.1736	Road freight transport	0.2178
Germany	0.1705	Motion pictures	0.2150

Czech Republic	0.1664	Logistics freight forwarding	0.2106
Netherlands	0.1600	Sound recording	0.2086
Latvia	0.1366	Distribution	0.1963

Table 1. Most and least restrictive countries and sectors in the STRI

Quantification of Trade Restrictiveness

Using the data from OECD STRI database, we can quantify the trade restrictiveness level for each country within each sector. To analyze cross-border data flow restrictions, we create a reduced dataset, which is a subset of the original data. This is accomplished through examining five specific policy measures concerning cross-border data flows. As described above, all policy measures are reported as answers to a series of *yes/no* questions. The restrictions used in this analysis are as follows:

- Free transfer of personal data or application of the accountability principle
- Transfer is possible only when certain private sector safeguards are in place
- Transfer is possible only to countries with substantially similar privacy protection laws or consent by government authority
- Fulfilling a combination of conditions is required before transfer is possible
- Transfer of personal data is prohibited

In this study, the dataset regarding service trade restrictiveness as a whole is referred to as the "full STRI." The dataset containing only data policy regulations is referred to as the "data STRI." These regulations provide a granular evaluation of data flow restrictions. By separating the dataset, we seek to understand how data restrictions are different from trade restrictions in general and the extent to which they overlap.

Clustering countries by Trade Restrictiveness

The STRI data poses a few unique challenges to clustering: its dimensionality and lack of a clear generative model make the data difficult to group. In this project, we use a pipelined model for clustering, first performing dimensionality reduction via principal component analysis, then using a variational Bayesian Gaussian mixture model for clustering.

Dimensionality Reduction

High-dimensional data is often difficult to cluster: with enough dimensions, everything begins to look like an outlier. The STRI dataset has 110 features (22 sectors, 5 policy areas per sector), but some features are not as "important" as others—a few features explain most of the variability in the data. This kind of data lends itself to dimensionality reduction, a set of algorithms that seek to find a lower-dimensional projection of the data that retains as much information as possible. One such approach is principal component analysis (PCA). Principal component analysis is a transformation that fits a set of orthogonal basis vectors to a series of data such that there is less variability across each axis than the previous one.

Mathematically, the principal component decomposition of a $p \times q$ data matrix $\mathbf{X}$ is given by $\mathbf{T} = \mathbf{XW}$, where $\mathbf{W}$ is a $p \times p$ matrix whose columns are the eigenvectors of $\mathbf{X}$. Each eigenvector corresponds to the "loading" of each component: it is the amount of each feature that is present in a given component, with the restriction that all eigenvectors must also be unit vectors. The first weight vector $\mathbf{w}_{(1)}$ is chosen such that the mean squared error (MSE) about that axis is maximized—in matrix form, this gives:

$$\mathbf{w}_{(\mathbf{1})} = \arg\max\left\{\frac{\mathbf{w}^T\mathbf{X}^T\mathbf{X}\mathbf{w}}{\mathbf{w}^T\mathbf{w}}\right\}$$

The remaining components are found by subtracting all previous components from $\mathbf{X}$:

$$\widehat{\mathbf{X}_k} = \mathbf{X} - \sum_{s=1}^{k-1} \mathbf{X}\mathbf{w}_{(s)}\mathbf{w}_{(s)}^T$$

Since all components are uncorrelated, any subset of components provides a projection of $\boldsymbol{X}$ into a lower-dimensional space, retaining a maximum of information. For this dataset, the first 3 principal components explained nearly 60% of the total variability and were the only ones retained in the clustering stage.

Clustering via Gaussian Mixture Models

This experiment looks to group countries with similar trade policies, regardless of what those policies may be. This is an unsupervised learning problem. The goal of unsupervised learning is to find intrinsic patterns in a dataset, rather than learning from feedback as in supervised learning problems. Specifically, we want to cluster the data, looking for ways to separate the data such that all the points in a given group are more similar to that group than to other groups. One typical approach to clustering is k-means clustering, which seeks to minimize the mean squared distance between points in each group and the group's centroid. This works well for some data, but it assumes that the data is roughly grouped into n-spheres, which is usually not the case.

Instead of using k-means, this project uses variational Bayesian inference to approximate the optimal parameters of a Dirichlet process Gaussian mixture model (GMM), as described by Blei and Jordan (2006). A mixture model is a more general unsupervised clustering algorithm than k-means which seeks to represent a dataset as a series of multivariate Gaussian distributions. This is particularly suited to datasets like the STRI, where the generative model for the data is unclear. Variational inference is used as an alternative to traditional Markov chain processes, and allows prohibitively slow stochastic processes to be turned into optimization problems. Applying this process to a mixture model assigns a different weight to each computed component, allowing unnecessary clusters to be absorbed into larger ones, resulting in a simpler model.

Results and Discussion

The variational inference used in our model retained 5 components from an initially 10-component model of the STRI data. The resulting 5 groups of countries each have unique characteristics, both for the "full STRI" dataset as a whole and for the "data STRI" dataset containing only data flow policies. This section outlines the results of both stages of the clustering pipeline, evaluates the characteristics of all groups, and explores the geopolitical factors that drive similarities in the data.

Overview

Figure 1 shows Scree plots for the first 10 components of PCAs of the "full STRI" and "data STRI" datasets. A Scree plot shows the variance explained by each principal component[1], as well as the cumulative variability explained from the first to the nth component.

The Scree plots show why PCA is effective for this dataset: for the full STRI dataset, nearly 60% of the total variability is explained in the first three components. This figure is even higher for the data STRI, with the first three components accounting for nearly 90% of the variability.

Table 2 provides an overview of all five groups for both the full and data STRI. The first column of each table lists the number of each group, which have been arbitrarily assigned from 1 to 5. The second column lists the number of countries in each group. The third column lists the mean composite STRI[2] for all countries in each group. The fourth column lists the mean deviation for each group. The mean deviation is a measure of the relative restrictiveness within each group. It is calculated by subtracting the mean composite STRI for each sector within a group from the mean composite STRI for each sector overall. Positive values indicate a relative tightening of restrictions, and negative values indicate a relative loosening.

[1] The variance explained by the nth principal component is the nth eigenvalue of the transformed covariance matrix.

[2] The composite STRI is the mean of the aggregate STRI values for all sectors in a given country.

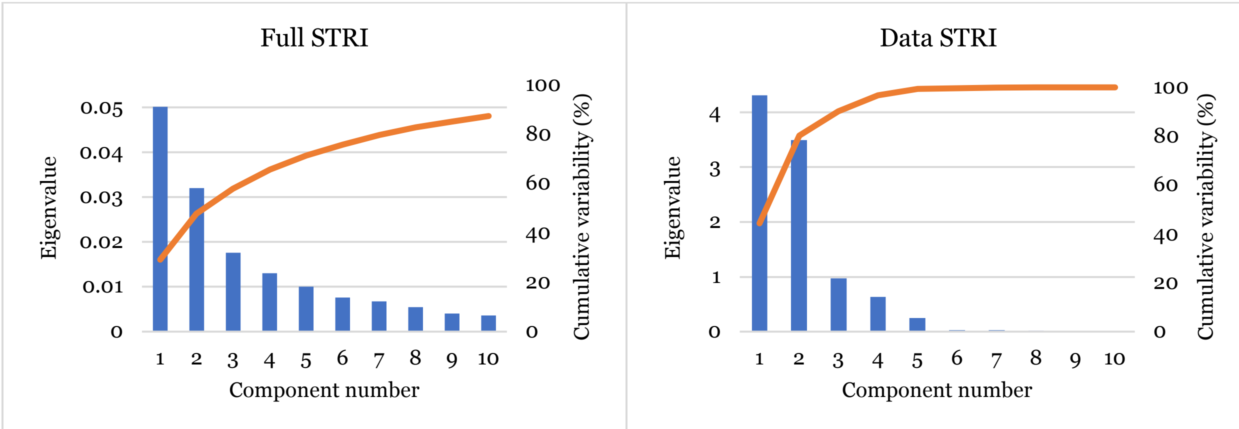

Figure 1: Scree plot of full (left) and data (right) STRI

In the full STRI, groups F1 and F4 are less restrictive than average, and groups F2, F3, and F5 are more restrictive. The mean squared error (MSE) for the groups in the "full STRI" is 0.0248, and the MSE for the "data STRI" groups is 0.0107.[3] This points to a higher dispersion among the groups for the full STRI: countries with similar sector STRI values are more likely to be grouped together in the "full dataset" than in the "data STRI" dataset.

Full STRI				Data STRI			
Group	**Count**	**Mean**	**Mean dev.**	**Group**	**Count**	**Mean**	**Mean dev.**
F1	24	0.2258	-0.0382	D1	29	0.2564	-0.0083
F2	7	0.3341	0.0728	D2	6	0.2601	-0.0035
F3	4	0.3441	0.0840	D3	6	0.2666	0.0045
F4	8	0.2485	-0.0209	D4	2	0.3059	0.0448
F5	1	0.4830	0.2272	D5	1	0.3954	0.1381

Table 2. Group statistics for full and data STRI

Group Analysis

Tables 3 and 4 list descriptions of all groups for both the full and data STRI datasets. The restrictiveness of each group is found in the second column. Each group has been labeled as either "liberal," "moderate," or "conservative," according to the mean deviation values from Table 2. Groups with a mean STRI below the global mean of 0.264 were considered to be liberal. Groups with a mean STRI near the mean were considered to be moderate, and groups with a mean STRI significantly above the global mean were taken to be conservative. Liberal economies favor free markets and openness of trade. Conservative economies, by contrast, favor privatization and often have a strong sense of nationalism. The defining features of each group are listed in italics at the beginning of the description column.

Each group is uniquely defined by its set of STRI values. In the full STRI, for example, Group F4 countries are especially restrictive in architectural and legal services. This can be explained by examining the distinguishing features of those sectors: in this case, both architecture and law are high-skilled fields, and Group F4 countries are generally restrictive to foreign administration of high-skilled services.

---

[3] MSE was calculated by summing the square distances between each sector in a group and the overall sector mean.

Full STRI		
Group	**Restrictiveness**	**Description**
F1	Liberal	*Open markets*: much less restrictive than average across all sectors
F2	Moderate	*Open borders*: fewer restrictions to movement of people
F3	Moderate	*State-owned monopolies*: market closure due to restricted foreign entry in transportation
F4	Liberal	*Domestic skilled labor*: more restrictive to foreign high-skilled laborers
F5	Conservative	*Closed markets*: much more restrictive than average across all sectors

Table 3: Group descriptions, full STRI

Data STRI		
Group	**Restrictiveness**	**Description**
D1	Liberal	*Mandated privacy*: trade partners must have similar privacy *regulations*
D2	Liberal	*Mandated safeguards*: private-sector firms must protect personal data
D3	Moderate	*Free transfer*: individual accountability for personal information
D4	Conservative	*Broad regulations*: multiple conditions for personal data transfer
D5	Conservative	*Closed borders*: transfer of personal data is prohibited

Table 4: Group descriptions, data STRI

Many trends seen in the groups become clearer from a geographic perspective: much of international trade policy is motivated by geopolitical concerns, and different regions often develop similar policies over time. Figure 2 shows a geographic projection of the policy groups, for both the full and data STRI.

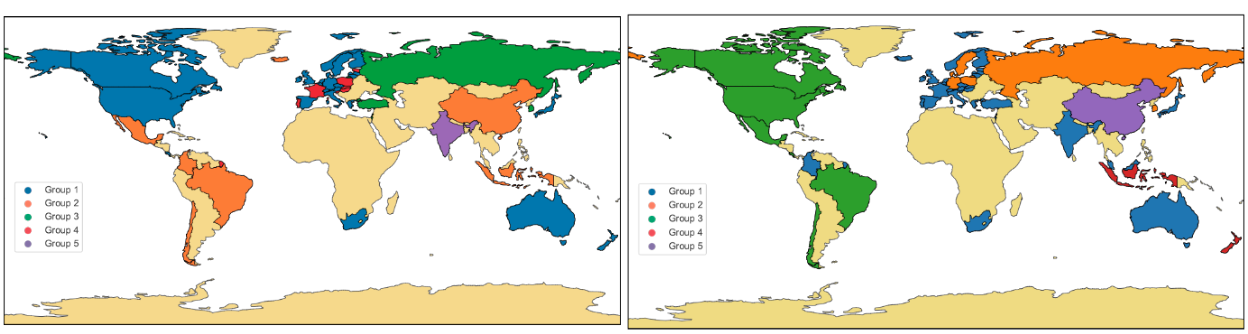

Figure 2: Geographic projection of policy groups, full (left) and data (right)

The most striking difference between the two projections is the division between the eastern and western hemispheres in the "data STRI." Many geopolitical factors contribute to this split: the adoption of the GDPR by many European countries has caused a homogenization in data policies for that region. Additionally, recent trade negotiations have prompted an agreement between the United States, Mexico, and Canada that standardizes regulations on IP transfer in the region ("United States-Mexico-Canada Agreement").

The reduced dataset reveals a disconnect between data trade restrictions and other trade restrictions in several countries. India, for example, is extremely restrictive and has the highest mean STRI across all sectors in the OECD database. In particular, high levels of restrictiveness in accounting (0.880), legal services (0.906) and rail freight transport (1.000) have driven an increased level of strictness. However, in the "data STRI" dataset, India belongs to a different category. Unlike similarly restrictive economies such as China, the Indian government allows cross-border transfer of personal data. The Indian Information Technology Act of 2000 outlined regulations that allow personal data to be transferred to foreign countries with similar regulations to India. However, in 2018, the Royal Bank of India issued a directive stating that all payment data held by payment companies should be held in local facilities. This reveals that, though India is working toward less restrictive in cross-border data flow relative trade policies, the inertia from restrictive trade policies in general services can still have notable impact in the digital trade environment.

Similarity Analysis

One challenge in this analysis is to quantify the difference in the separation of groups between the full and data STRI. This can be evaluated using Jaccard indices, as described by Hwang and Hung (2018). The Jaccard similarity index, also known as the intersection-over-union coefficient, is defined as the ratio between the size of the intersection and the size of the union of two sets:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

Computing the 5×5 matrix of the Jaccard coefficients between the full and data STRI allows us to quantify the overlap between the two. The coefficients can be seen in Table 5, with values near 1 indicating more complete overlap and values near 0 indicating little overlap.[4] Countries with open markets (Group F1) tend to value data privacy (Group D1); countries with open borders (Group F2) tend to promote accountability in data transfer (Group D3); and countries with several state-owned monopolies (Group F3) tend to mandate private-sector safeguards for personal data (Group D2). On the other hand, Group D4 and D5, with conservative cross-border data flow restrictions, actually with overlap with Group F1 and Group F2, which are countries with open markets or borders. This reveals that the countries with less restrictive in general service trading, are not necessary to continue such policies in the digital trading environment.

Group (data)	**Group (full)**				
	F1	F2	F3	F4	F5
D1	**0.472**	0.059	0.065	0.233	0.034
D2	0.111	-	**0.250**	0.077	-
D3	0.111	**0.300**	-	-	-
D4	0.040	0.125	-	-	-
D5	-	0.143	-	-	-

Table 5: Group Similarity Coefficients between full and data STRI

Trend Analysis

Figure 3 shows trends in mean STRI per group, using a time series constructed with OECD data from 2014-2018. The columns indicate the percentage change in mean STRI from one year earlier, with positive values indicating a trend toward stricter regulations and negative values indicating a trend toward open markets.

[4] Values of 0 are shown as "-" for clarity.

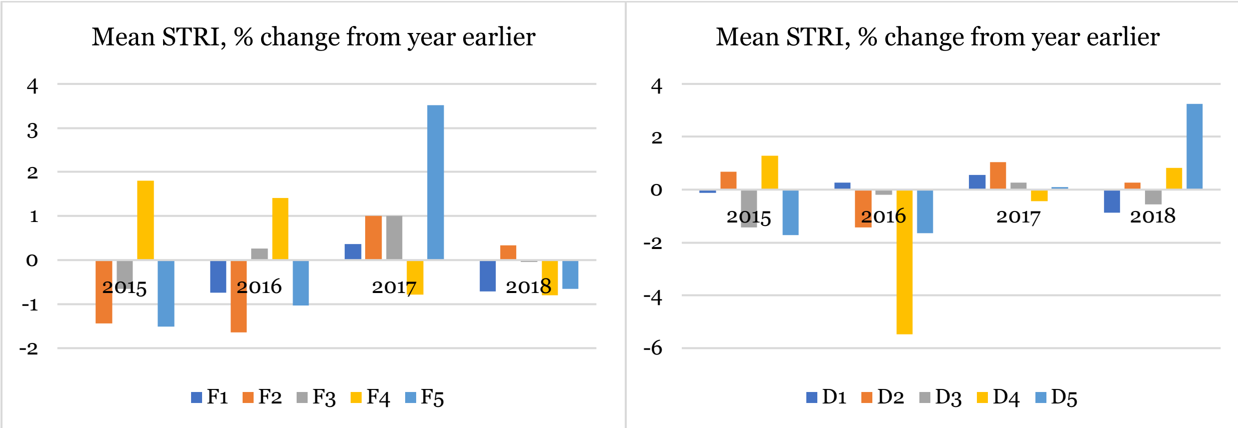

Figure 3: Mean change from year earlier, full (left) and data (right) STRI

Increased levels of restriction in the full STRI groups have been highly motivated by efforts to localize supply chains, especially in 2017. Group F5, for example, saw a significant tightening of restrictions in 2017 due to "Rule 153" in India. The new regulation shifted toward explicit preferences for local suppliers, authorizing the government to "provide for mandatory procurement of any goods or services from any category of bidders."

Data localization drives trends in the data STRI groups. In Turkey, regulations were passed in 2015 requiring that "operations and data related to e-payment and e-money services must be located in Turkey." China passed the Cybersecurity Law of the PRC in 2017 that required "'Personal information and important data' collected and generated in China by critical information infrastructure operators to be stored domestically." This change can be seen to account for the rapid increase in mean STRI for group D5.

Conclusion

This paper presents a quantitative framework for categorizing cross-border data flow trade policies and analyzing the ways they differ from other service trade restrictions in general. A mixture-based clustering pipeline was used to group the OECD STRI database into clusters that share common traits. A dataset specific for cross-border data flow policies was constructed to measure the differences between data-related policies and trade policies in service as a whole; multiple analyses were conducted on both datasets to assess the trends within each and the extent to which they differ.

Our analysis reveals the following: First, given the STRI data, the localization effect is stronger among cybersecurity restrictions. This is substantiated by a lower mean deviation and lower geographic dispersion among the restricted STRI policy groups. Second, while there is a significant overlap between cross-border data flow restrictions and other trade regulations in liberal and moderate policy groups, there is little overlap between the two for highly restrictive, conservative policy measures.

The preliminary results from this study highlight that in the digital trade environment, trade policies related to cross-border data flow do not necessarily follow the same patterns as service trade in general. Though some countries will have a less restrictive data trade policy as a continuation of their openness within the general service trade, some turn to implement more restrictive data trade policies for digital trade. Countries which implement open data trade policy may swing if they have restrictive service trading policies in general. While the trade policies in services are covered by the WTO General Agreement on Trade in Services (GATS), there is no global policy schema for cross-border data flows within digital trade. The differences this study reveal indicate the need for new trade policy norms in digital trade.

We acknowledge two key limitations with this study. First, the data contained in the STRI is somewhat limited, especially with regards to cross-border data flow restrictions. Future research could address this issue by gathering more data on IP trade policy and cybersecurity restrictions and incorporating this into the STRI data. Second, due to the stochastic nature of our model, the groupings are highly sensitive to exogenous parameters and can vary slightly between successive evaluations. This effect could be mitigated by using a Monte Carlo process to repeat the clustering process with slightly different parameters each iteration until a stable solution is reached.

As the part of the roadmap for this project to promote the norms related to cyberspace within international digital trading, we will further examine the evolution of the data restrictiveness to understand the dynamics of trade policy within the digital trade environment.

Acknowledgements

This work is partially supported by the MIT Internet Policy Research Initiative (MIT IPRI), Cybersecurity at MIT Sloan (the MIT Interdisciplinary Consortium for Improving Critical Infrastructure Cybersecurity, MIT-(IC)3) and MIT Policy Lab at the Center for International Studies (MIT IPL).

REFERENCES

Attias, H. 2000. "A Variational Bayesian Framework for Graphical Models," *Advances in Neural Information Processing Systems* (12), pp. 209–215.

Benz, S. 2017. "Services trade costs," *OECD Trade Policy Papers* (200) (doi: 10.1787/dc607ce6-en).

Bishop, C. M. 2016. *Pattern Recognition and Machine Learning*, New York, NY: Springer New York.

Blei, D. M., and Jordan, M. I. 2006. "Variational inference for Dirichlet process mixtures," *Bayesian Analysis* (1:1), pp. 121–143 (doi: 10.1214/06-ba104).

Borchert, I., Gootiiz, B., and Mattoo, A. 2012. "Guide to the Services Trade Restrictions Database," *Policy Research Working Papers* (doi: 10.1596/1813-9450-6108).

Cohen, B., Hall, B., and Wood, C. 2017. "Data Localization Laws and Their Impact on Privacy, Data Security and the Global Economy," *Antitrust Magazine* (32:1), pp. 107–114.

Chander, A., and Le, U. P. 2014. "Breaking the Web: Data Localization vs. The Global Internet," *SSRN Electronic Journal* (doi: 10.2139/ssrn.2407858).

Ferencz, J. 2019. "The OECD Digital Services Trade Restrictiveness Index," *OECD Trade Policy Papers* (doi: 10.1787/16ed2d78-en).

Gervais, A. 2017. "Estimating the impact of country-level policy restrictions on services trade," *Review of International Economics* (26:4), pp. 743–767 (doi: 10.1111/roie.12340).

Grosso, M.G., Gonzales, F., and Miroudot, S. 2015. "Services Trade Restrictiveness Index (STRI): Scoring and Weighting Methodology," *OECD Trade Policy Papers* (200) (doi: 10.1787/18166873).

Hwang, C.-M., Yang, M.-S., and Hung, W.-L. 2018. "New similarity measures of intuitionistic fuzzy sets based on the Jaccard index with its application to clustering," *International Journal of Intelligent Systems* (33:8), pp. 1672–1688 (doi: 10.1002/int.21990).

Koopman, R., and Maurer, A. 2018. *World Trade Report 2018*, Geneva, Switzerland: World Trade Organization.

Madnick, S., Johnson, S., and Huang, K. 2019. "What Countries and Companies Can Do When Trade and Cybersecurity Overlap," *Harvard Business Review*, Harvard Business Publishing, January 4.

Nordås, H. K., and Ragoussis, A. 2009. "Methodology for deriving the STRI," *OECD Experts Meeting on the Services Trade Restrictiveness Index (STRI)*.

Ryan, P. S., Falvey, S., and Merchant, R. 2013. "When the Cloud Goes Local: The Global Problem with Data Localization," *Computer* (46:12), pp. 54–59.

Selby, J. 2017. "Data localization laws: trade barriers or legitimate responses to cybersecurity risks, or both?," *International Journal of Law and Information Technology* (25:3), pp. 213–232 (doi: 10.1093/ijlit/eax010).

"Services Trade Restrictiveness Index (STRI): Scoring and Weighting Methodology." 2015. *OECD Trade Policy Papers* (doi: 10.1787/5js7n8wbtk9r-en).

"United States-Mexico-Canada Agreement." (n.d.). *Office of the United States Trade Representative* (available at https://ustr.gov/trade-agreements/free-trade-agreements/united-states-mexico-canada-agreement; retrieved February 28, 2019).

Wold, S., Esbensen, K., and Geladi, P. 1987. "Principal Component Analysis," *Chemometrics and Intelligent Laboratory Systems* (2), pp. 37–52.

"World Trade Statistical Review 2018." 2018. *World Trade Statistical Review* (doi: 10.30875/0ab3aa40-en).