United Nations University
Institute on Comparative Regional Integration Studies

WORKINGPAPER
No. 02, 2023

Towards an Exhaustive Regional Integration Agreements Database

Justine Miller and Samuel Standaert

UNU CRIS

About the authors

Justine Miller is a PhD Fellow at UNU-CRIS and Ghent University. She is working towards understanding the concept of the "Spaghetti bowl" of regional integration under the supervision of Samuel Standaert. She aims to quantify the phenomenon's impact on countries by using network analysis tools and counterfactual techniques. Justine holds a Master's Degree in Advanced Studies in Economics from the KULeuven and one in Economic Geography from Utrecht University. Aside from her studies, Justine has worked at the Dutch Central Bank, where she conducted research on financial integration through the lens of network analysis. She also contributed to research projects in both innovation studies and welfare economics at Utrecht University.

Contact: jmiller@cris.unu.edu

Samuel Standaert is a Professorial Fellow and coordinator of the Regional Integration Knowledge System cluster at UNU-CRIS and Assistant Professor at Ghent University. In this role, he works on various topics, including the effects and determinants of regional integration agreements; the gravity model of international trade; sustainable development; Bayesian econometrics; and migration policy. He also coordinates (with Professor Glenn Rayp) the work at UNU-CRIS on RIKS 2.0. At Ghent University, he supervises PhD students working on various projects in the field of regional integration agreements and teaches several courses on economics.

Contact: sstandaert@cris.unu.edu

About UNU-CRIS

The United Nations University Institute on Comparative Regional Integration Studies (UNU-CRIS) is a research and training institute of the United Nations University, a global network engaged in research and capacity development to support the universal goals of the United Nations and generate new knowledge and ideas. Based in Bruges, UNU-CRIS focuses on the provision of global and regional public goods, and on processes and consequences of intra- and inter-regional integration. The Institute aims to generate policy-relevant knowledge about new patterns of governance and cooperation and build capacity on a global and regional level. UNU-CRIS acts as a resource for the United Nations system, with strong links to other United Nations bodies dealing with the provision and management of international and regional public goods.

The mission of UNU-CRIS is to contribute to generate policy-relevant knowledge about new forms of governance and cooperation on the regional and global level, about patterns of collective action and decision-making.

UNU-CRIS focuses on issues of imminent concern to the United Nations, such as the 2030 Development Agenda and the challenges arising from new and evolving peace, security, economic and environmental developments regionally and globally. On these issues, the Institute will develop solutions based on research on new patterns of collective action and regional and global governance. The Institute endeavours to pair academic excellence with policy-relevant research in these domains.

For more information, please visit www.cris.unu.edu

in alliance with

VUB VRIJE UNIVERSITEIT BRUSSEL

UNIVERSITEIT GENT

Vlaanderen verbeelding werkt

BRUGGE

Abstract

Recent decades have seen regional integration agreements (RIAs) become a dominant feature of international relations. The rapid proliferation of agreements has sparked the need to compile information on the phenomenon. In answer to this call, several projects started gathering data on trade agreements and regional organisations and have become widely used in regional integration studies. Despite these efforts to collect data, some challenges remain. One issue is that no single database offers a complete picture of the current world trading scene owing to data collection and coding limitations. Additionally, when the data sets overlap, they are often inconsistent and sometimes provide conflicting information that could potentially bias analyses.

This paper provides an overview of the five most prominent databases on RIAs. We show how these databases can be combined to create a more comprehensive repository of trade agreements and regional organisations. The final database contains 1,149 trade agreements and regional organisations from the early 1910s to 2020. The database contains information on treaty characteristics, including the year of signature, entry into force, or inactivity of an agreement. We also provide categorisations of the type and region of RIAs. Moreover, we track the changes in membership: following late accessions or early withdrawals to RIAs. We provide this detailed country membership data in both panel and dyadic formats, adjusted for the year of entry into force when available. By linking each treaty to a parent RIA, we can easily study how the content of agreements evolves.

Keywords:
Trade agreements, regional organisation, international trade

Table of Contents

List of Abbreviations

ASEAN	Association of South East Asian Nations
CAN	Andean Community
CAFTA	Central American Free Trade Area
CAFTA-DR	Central American Free Trade Agreement Dominican Republic
CEFTA	Central Europe Free Trade Agreement
CROP	Comparative Regional Organizations Project
DESTA	Design of Trade Agreements
EAP	East Asia and the Pacific
ECA	Europe and Central Asia
EC	European Community
EFTA	European Free Trade Association
EIA	Economic Integration Agreement
EU	European Union
FTA	Free Trade Agreement
FYROM	Former Yugoslav Republic Of Macedonia
GATS	General Agreement on Trade in Services
GATT	General Agreement on Tariffs and Trade
GPTAD	Global Preferential Trade Agreements Database
LAC	Latin America and the Caribbean
MENA	Middle-East and North Africa
MSG	Melanesian Spearhead Group
NA	North America
NAFTA	North American Free Trade Organisation
PLO	Palestine Liberation Organisation
PTA	Preferential Trade Agreement
RIA	Regional Integration Agreement
RIKS	Regional Integration Knowledge System
RO	Regional Organisation
RTA	Regional Trade Agreement
SA	South Asia
SACU	Southern African Customs Union
SADC	Southern African Development Community
SSA	Sub-Saharan Africa
TA	Trade Agreement
UNMIK	United Nations Interim Administration Mission in Kosovo
UNU-CRIS	United Nations University Institute on Comparative Regional Integration Studies
WB	World Bank
WTO	World Trade Organisation

1. Introduction

The last three decades have witnessed a proliferation of regional integration agreements (RIAs) promoting reciprocal preferential trade liberalisation, i.e., barriers to trade are reduced but only between the member states.[1] This growth has given rise to a broad stream of literature that studies everything about these RIAs, from the determinants of their formation to their effectiveness at fostering trade (e.g., Baier et al., 2014; Baier et al., 2019; Limão, 2016). These studies require vast and precise amounts of data covering not only the identifying variables of the agreements but also the structure, content, and membership information of these RIAs. Over the years, several projects have been developed to address these needs.

Despite the quality and carefulness of these projects, they often seem to disagree. First, differences in data collection processes imply that the databases cover distinct types of RIA. For example, the World Trade Organisation (WTO) focuses on *WTO-notified* treaties, which excludes about 55% of RIA treaties. Similarly, other databases on RIAs end up missing between 25% and 70% of RIAs (cf. infra). Second, for those RIAs that appear in multiple databases, we often observe discrepancies in the treaty characteristics, including dates and membership, as is the case for 13% of treaty dates (cf. infra).

These two sources of divergence can affect the results of studies that build on this data. For instance, as we will show in this paper, estimations of the impact of trade agreements on trade flows may vary drastically depending on the database that is used. The reason is that the estimations compare trade flows between countries with and without an RIA. If those two categories are misidentified, the estimated impact of an RIA will be biased. Accordingly, it is crucial to construct a database that is as exhaustive as possible and that provides cohesive information.

This paper identifies the discrepancies between five existing databases on RIAs. We then construct a combined database that resolves the differences and brings together the information from all five databases. This combined data aims to provide an exhaustive list of past and current RIAs worldwide from the start of the twentieth century onwards. We consider all agreements that preferentially eliminate the barriers to trade in goods or services. Specifically, we consider Trade Agreements (TAs) and Regional Organisations (ROs). The former focuses on liberalising the international flow of goods, services, or both. The latter are organisations that set up an institutional superstructure (supranational or intergovernmental). While a TA will typically consist of a single treaty signed between the partner states, many ROs have a founding treaty and subsequent amending treaties.
In total, we bring together data from six databases. Four that focus predominantly but not exclusively on TAs, and two databases that solely focus on ROs and their founding and amending treaties. These databases are:[2]

- DESTA: The Design of Trade Agreements Database (Dür et al., 2014)
- WTO: World Trade Organisation Regional Trade Agreements Database (World Trade Organisation, n.d.)
- WB: The Global Preferential Trade Agreements Database (World Bank, n.d.)
- RIKS: The Regional Integration Knowledge System (UNU-CRIS, n.d.)
- CROP: The Comparative Regional Organizations Project (Jetschke et al., 2021)

This combined data contains information on 1,149 treaties. For each treaty, it lists membership, dates (i.e., the year of signature, entry into effect), and other basic descriptives. It also brings together any additional information listed in the source data.

While we discuss differences and errors between the databases throughout this paper, we aim to highlight their respective qualities and contents. These differences between the data sources are a testament to the inherent difficulty of tracking the membership and content of RIAs over the past century.

[1] These agreements have been given many different names over the years. The WTO uses Regional Trade Agreements (RTAs) as the collective noun, but it is common to see Economic Integration Agreements (EIAs), Preferential Trade Agreements (PTAs), or even FTA+. Here we use Regional Integration Agreements even though not all agreements are signed on a regional basis. Given that the definition of regions is loosely defined, we solely work with the term RIAs for consistency's sake.

[2] We focus solely on databases at the global level and not at the regional level.

The rest of this paper proceeds as follows. The next section provides a brief overview of the most often used datasets, including those we use to build our dataset. Section 3 discusses how the data is merged and provides examples of the dissimilarities encountered. Finally, Section 4 provides some evidence of how analyses of RIAs can be affected by the differences in the databases, including gravity models of international trade and network analysis tools.

2. The Databases

This section presents an overview of the data sets covering trade agreements and regional organisations. In particular, we expand on the goals of the databases, their scope, the type of information they capture, the methods used to collect the data, and how regularly the data is updated.

2.1 World Trade Organisation Regional Trade Agreements Database

The World Trade Organization's Regional Trade Agreements (RTA) Database provides information on all treaties notified to the WTO. The database is a repository for the legal texts, annexes, and related documents of all agreements notified to the WTO. It provides information on the content of provisions, the parties involved, the type of the agreement, as well as the status of implementation of a treaty.

Not all RTAs have to be notified to the WTO, which explains the limitations of the database. In particular, WTO members are required to notify the WTO when a treaty is signed, as specified by Article XXIV of the General Agreement on Tariffs and Trade (GATT) and by Article V of the General Agreement on Trade in Services (GATS). The notification obligation does not hold if at least one party is not a WTO member, even though countries are strongly encouraged to notify treaties with the WTO. In practice, agreements for which not all parties are WTO members experience less scrutiny, particularly regarding changes to already existing agreements.

The RTA database was developed in 2006 as part of a push towards greater transparency of trade policies. Concretely, every time a treaty – founding or amending – is notified to the WTO, it is added to the database. The database is continuously maintained by the RTA Section of the WTO Trade Policies Review Division.

2.2 Design of Trade Agreements

The DESTA project compiles data on trade agreements that go beyond those notified to the WTO, and captures the content of the agreements (Dür et al., 2014). The database includes free trade agreements, partial trade agreements, services agreements, and customs unions. Dür et al. (2014) develop a measure of the depth of integration,[3] an index that ranges from 0 to 7 and increases in increments of one, by counting the number of substantive provisions across tariff reductions, services trade, investments, standards, public procurement, competition and intellectual property rights provisions (Dür et al., 2014). They define a substantive provision as, for instance, a national treatment clause in the services chapter.

For the data collection process, the authors search websites and documents of foreign ministries and governmental institutions. In its original release, DESTA contained 733 treaties covering the period 1945-2009, and the authors found that most of the agreements were not notified to the WTO (Dür et al., 2014). Independent coders and a referee were enlisted to manually code the contents of the agreements, with 587 of the 733 agreements analysed in detail and categorized based on provisions ranging from market access to dispute settlement.

The data set was originally published in 2009, covering the period from 1945 to 2009, and has since been updated regularly to cover the period from 1948 to 2019.[4]

[3] The depth of a TA refers to the number of provision themes it covers and the level of cooperation agreed upon by member countries. Accordingly, two countries can be linked via multiple TAs of varying depths of integration. A deep relationship between two countries can also result from (the combination of) more than one (shallow) agreement that covers various trade-related topics such as services, investments, taxes, etc.

[4] The latest update was in January 2022.

2.3 Regional Integration Knowledge System

The Regional Integration Knowledge System (RIKS) has been developed by the United Nations University Institute on Comparative Regional Integration Studies (UNU-CRIS) since 2016, with earlier versions starting in the early 2000s. The platform provides public access to detailed (historical) membership information on regional organisations worldwide and tracks the changes in the level of integration of these organisations (UNU-CRIS, n.d.). This information is provided on a yearly basis for each organisation and for each country.

RIKS distinguishes between the year of the signing of the agreement (and withdrawal if applicable), the year of ratification, and the year of entry into effect at the country level. When available, the data also indicates whether a country was suspended and how long that suspension lasted. Moreover, it contains various indicators about trade, migration, and foreign direct investment (among others) since 1960, subject to data availability.

To be included in the database, organisations have to fulfil certain criteria: (1) Member states have to belong to the same geographical sub-region or neighbouring geographical sub-regions; (2) There have to be at least three member states; (3) Regional integration has to be explicitly set among the purposes in the establishment of the organisation/arrangement, as defined in treaty texts. The current data set contains information on 155 organisations. To compile this data, the team examines several data sources in a preferential order: (1) the official website of the organisations; (2) the treaty texts (the signatory states mentioned); (3) the Wikipedia page of the organisations; (4) the Yearbook of International Organisations; (5) press releases, newspaper articles, books.

2.4 Comparative Regional Organisations Project

The CROP project aims to comprehensively gather information on the design of regional organisations globally. Its objective is to better understand the factors that shape institutional design choices, their outcomes, and how they spread (Jetschke et al., 2021). Specifically, the project studies ROs that aim to both promote economic prosperity and development between its member states, and promote the peace in the member states' respective regional environments (Jetschke et al., 2021). The authors selected the ROs according to three criteria: (1) the RO must consist of at least two contiguous states who define their membership therein on a regional basis (their definition of region is somewhat more flexible to allow for transcontinental organisations to be considered); (2) the RO must be multipurpose, rather than focusing on a single issue area; (3) the RO must be sufficiently institutionalised to require regular meetings and rules governing decision making (Jetschke et al., 2021). This selection process resulted in 80 organisations, covering 276 founding and amending treaties.

In their original paper, the authors constructed this database to compare organisations across time as captured by their founding treaties, treaty revisions and amendments between 1945 and 2018 (Jetschke et al., 2021). The main contribution of the data lies in coding the policy areas in which member countries collaborate, whereas most data sets with a similar scope tend to focus either on specific areas or on identifying the degree of institutionalisation of organisations (Jetschke et al., 2021).

The authors created a codebook consisting of fifteen policy areas. They provided training to manual coders and imposed quotation requirements to improve accuracy and reproducibility (Jetschke et al., 2021). The coders were expected to go through over 300 questions and to mark the response items as 1 if affirmative and 0 if absent. Accordingly, this approach makes CROP comparable to DESTA in that both provide detailed coding of the content of agreements and organisations.

To our knowledge, the CROP data is no longer being updated.

2.5 The Global Preferential Trade Agreements Database

The Global Preferential Trade Agreements Database (GPTAD) is developed jointly by the World Bank and the Center for International Business at the Tuck School of Business at Dartmouth College (World Bank, n.d.). The database contains the original text of bilateral, plurilateral, and regional trade agreements, including free trade agreements and customs unions signed since post-World War II. Moreover, the database includes agreements that have been and have not been notified to the WTO.

The type of data collected includes information on the parties involved, the scope of the agreement, and the tariffs applied to member countries. The data is gathered through official government sources, RIA texts, and other relevant documents. The database is updated regularly and currently comprises over 300 treaties.

2.6 Additional Databases (Not Included in the Merging)

There are four other databases that are often used in the analyses of RIA. However, because they are built on the datasets described above, all relevant variables are covered using the other databases, mainly that of the WTO. We briefly expand on their characteristics below.

First, there is Mario Larch's Regional Trade Agreements Database (Egger and Larch, 2008). It includes all RIAs, including Customs Unions, Partial Scope and Economic Partnership Agreements, that were notified to the World Trade Organization for the last 70 years from 1950 to 2022 (570 agreements). Accordingly, it directly builds on the WTO database. The authors also provide data on country membership and treaty dates, such as the year of signature and year of entry into force. The database was last updated in 2022.

A second database is compiled by Hofmann et al. (2017). The authors also focus on treaties notified to the WTO, specifically on those that were active at the time. Hofmann et al. (2017) analyse how the content of treaties has changed over time by studying the type of provisions contained in treaties. They consider both WTO+ provisions, which fall within the WTO regulations, and WTO-X provisions, which go beyond the WTO mandate. Accordingly, they study the so-called "horizontal depth" of trade agreements, since they can only look at the coverage of provisions of treaties, rather than ranking the provisions based on the extent to which they further the level of integration (this would be then be referred to as "vertical depth"). Moreover, they also construct three indexes, one that counts the total number of provisions included in a treaty, one that only focuses on economic provisions, and one constructed from a principal component analysis conducted on the entire set of provisions. As far as we are aware, their has not been updated since.

The third database we considered was designed by Bergstrand and Baier in 2004, in partnership with the NSF-Kellogg Institute. It covers all trade agreements between 195 countries during the period 1950-2017 (NSF-Kellogg Institute, n.d.). The database further specifies the type of agreement (free trade agreement, economic union, ...) and provides pdf files of the original treaties or documents related to it. The data was last updated in 2021 and covers all treaties up to 2017.

The final database we examined is the Heterogeneity of Trade Agreements Database. It was constructed by Kohl et al. (2016) who aimed to study both the coverage of trade provisions and their enforceability. Their data builds on the GPTAD (WB) and includes 296 agreements (193 WTO-notified), covering the period 1948-2011. Concretely, Kohl et al. use the WB database to assign a binary variable to each policy domain covered by the RIA under investigation. Then they study the degree of legal enforceability by studying keywords and commitments based on the type of provision. Ultimately, their additional depth variables differ in two main ways from the ones of Dür et al. (2014): (1) they have different coverage of trade agreements, and (2) Kohl et al. (2016) also specifically study the enforceability of provisions. Moreover, Kohl et al. (2016) directly build on the methodology of other articles (see Horn et al., 2010; Orefice & Rocha, 2014) whereas DESTA is less comparable to other works. As is the case for the Hofmann et al. (2017) data, it has not been updated since the initial release.

3. Methodology

The five data sets are merged using a three-step process:

1. Manual merging of treaties based on their names and year of signature.
2. Automated extraction and harmonisation of treaty characteristics.
3. Creation of files capturing country membership information over time using the exhaustive list of treaties constructed in steps 1 and 2.

3.1 Combined Database

3.1.1. Manual Merging and Harmonisation of the Data

Manual Merging

The first step of the merging consists in manually matching the treaties across the databases. This step had to be conducted manually as unpredictable discrepancies in treaty names, treaty dates, and the coverage of treaties between the databases prevent an automated matching.

Our merging procedure aims to be comprehensive and includes as many treaties as possible. Specifically, we include agreements that cover trade in goods, services, or both. Whenever an agreement encompasses goods and services, we prioritise the information from the goods portion as this tends to be more comprehensive. RIAs that have not come into force before or during 2020 are excluded from our analysis. Additionally, we only include RIAs for which we have complete membership information. If the year of entry into force is not specified, we still include the treaty in our analysis.

The datasets were matched two-by-two. To identify a match, we first compare the RIA name and its date of signature. If the treaty could not be matched this way, we compare more variables, such as the year of entry into force or different naming conventions.

While the treaty name typically results in an unambiguous match, the databases tend to adopt multiple and often highly similar acronyms that necessitate thorough vetting. Table 1 lists the most prevalent examples, including the number of treaties affected. While none of these issues is irreconcilable, they limit the potential for automation of the merging procedure between the databases and significantly increase the chance of overlooking a match between databases.

Name Issue	Prevalence	Percentage Prevalence
Country order	177 treaties	15.4%
EU-EC	110 treaties	9.6%
Country name/recognition	19 treaties	1.7%
Serbia and Montenegro	18 treaties	1.6%
North Macedonia reference	17 treaties	1.5%
Northern Triangle, Central America	14 treaties	1.2%

Table 1. Examples of different naming conventions across the databases.

The most predominant difference lies in the order in which the countries are listed in the RIA name. While DESTA tends to order the countries alphabetically, the WTO and the WB follow a different pattern (that also differs between the two). Another difference arises from the fact that DESTA refers to the European Union as EC, which itself stands for European Community, but the WTO refers to it as the EU or the EC based on the year of signature of the agreement.

Moreover, the status of independence and country recognition also influences the databases. For instance, most databases refer to Taiwan, but the WTO uses Chinese Taipei. The WB only refers to China in the main name but mentions Taiwan in the comments. Similarly, the WB and the WTO use both Kosovo and United Nations Interim Administration Mission in Kosovo (UNMIK). Finally, DESTA solely uses PLO (Palestine Liberation Organization), WTO uses Palestine and the WB alternates. Likewise, countries that have separated or become independent during the existence of an agreement are encoded differently

across databases. Serbia and Montenegro perfectly illustrate such a case. Between 1992 and 2003, the Republics of Serbia and Montenegro were both part of Yugoslavia. In 2003, Serbia and Montenegro became a political union until Montenegro separated in 2006, and both countries became independent. Accordingly, only one agreement (signed in 1980 with the EU) mentions Yugoslavia. The agreements signed between 1996 and 2002 in WTO and DESTA involve a combination of all nomenclatures. DESTA and the WB start referring to Serbia and Montenegro separately as early as 2003 (except for one treaty).[5] The other databases use the political union until 2006, from which point they use the independent countries.

Similarly, the databases refer to Macedonia in different, sometimes internally inconsistent, ways. In particular, the WB refers to Macedonia either as Macedonia or as the Former Yugoslav Republic of Macedonia (FYROM). DESTA consistently refers to it as Macedonia, and WTO uses the country's current name: North Macedonia.

Finally, the encoding of specific treaties can differ, with an agreement appearing once in one database and multiple times in another. That is the case for the agreement between Mexico and the countries of the Northern Triangle. All databases record the agreement once, but the WTO records three separate treaties for each country of the Northern Triangle (Guatemala, Honduras and El Salvador). To avoid triple counting the same agreement, we match one of the treaties from the WTO with the others and discard the others from the merged version.[6] Similarly, agreements involving Central American countries also appear multiple times in the WTO but not in the other databases.

In Appendix A, we list additional cases for which no clear pattern could be discerned and expand on whether or not we merge the treaties.

Aside to the treaty name, we also base our merging on the treaty dates. Most databases discern between the year of signature of an agreement, and when available, its year of entry into force and year of inactivity/end date. It is important to distinguish between the year of signature and the year of entry into force since the signature of a treaty only serves as a signal of commitment to implement a treaty rather than securing the actual implementation. A treaty enters into force once enough members have ratified. The ratification process consists in the member countries consenting to the treaty at a national level. Some treaties require all members to ratify nationally, but others only require a certain fraction of the members to ratify it for the treaty to become binding. Treaties can sometimes remain unratified for many years, or never come into force at all. Unfortunately, the date of entry into force is not always systematically reported, so some treaties might effectively be implemented even though the databases do not report it to be. In our data, we have a year of entry into force for over 65% of the treaties, and on average, the ratification takes place rather fast after a treaty is signed (over one year and four months).

Those three types of treaty dates are the variables for which there is the most heterogeneity across the databases, and often required manual intervention. The Southern African Customs Union (SACU), the oldest customs union in the world, is a worthy example of those large discrepancies. It was first established in 1910 this treaty, with two major amendments following in 1969 and 2002, as stated on SACU's official website.[7] RIKS is the only database that marks the official start of SACU. DESTA has treaties only for these later two agreements, but not for the establishment of the customs union in 1910. CROP does not seem to contain any treaty for it at all. The first SACU treaty in both the WB and WTO databases is only in 2002, the year the last main treaty of SACU was signed. It is matched with the DESTA equivalent. Our final merged database includes three treaties for each of these significant agreements. In their article, Egger & Larch (2008) also noted that treaties such as SACU illustrated issues arising from earlier treaties not being notified to the WTO.

In 21 instances, there are different dates of signature between the databases, but the same year of entry into force. Such is the case for the agreement between Canada and Panama for which DESTA has 2009 as the year of signature and 2013 as the year of entry into force, whereas in the WTO the date of signature is 2010.

[5] Macedonia-Serbia and Montenegro, signed in 2005.

[6] DESTA marks the WTO treaty as the one matched with Honduras, so we follow the same approach.

[7] Retrieved from https://www.sacu.int/show.php?id=394 and last accessed on 03/02/2023.

Moreover, the databases do not consistently use the founding treaties of regional organisations as its main identifier. Take the case of the Andean Community (CAN). RIKS, DESTA, and CROP all include the RO from 1969. However, RIKS uses the term Andean Community, while DESTA and CROP both refer to it as the Cartagena Agreement. The first time CAN appears in WTO is in 1987, which corresponds to the second CAN treaty, marked by the ratification of the Quito Protocol. CAN only makes it first appearance in WB in 1996 with the ratification of the Protocol of Trujillo.

Harmonisation of the Data

After matching the treaties, we harmonise the information across the common variables in all the databases. In particular, we reconcile data on the treaty names and treaty dates. For regional organisations, we prioritise the names from RIKS when available and CROP otherwise. That is because RIKS focuses more on founding treaties, so the names are more easily identifiable. For trade agreements, we prioritise the names from DESTA as it is the largest database. For the other treaties, we use WTO first, followed by WB and CROP's names for the remaining ones.[8]

We follow the same order to harmonise the data on the three types of treaty dates. All databases have information on the date of signature of a RIA. All databases on trade agreements (DESTA, WTO, WB) have information on the date of entry into force. For regional organisations, RIKS only has this information at the country level, but not at the treaty level. CROP does not have either. The date of inactivity/end of a RIA are solely found in RIKS and WTO. To illustrate the inconsistencies in treaty dates between the databases, we plot the differences between the original databases and our final harmonised version in Figure 1.

In the majority of cases, the databases have a delay compared to the combined value. Moreover, a few treaties have differences larger than 20 years. For instance, the Central American Free Trade Area (CAFTA) is one of the rare ones that has a year of inactivity that differs by over 40 years. The WTO marks the treaty as ending only one year after its signature (1961). However, RIKS has an ending date in 2004, when the treaty became replaced by its successor, the Central American Free Trade Area (CAFTA) Dominican Republic treaty. Nevertheless, DESTA, RIKS, and WTO all have a similar date of signature.

For 78 treaties (nearly 7% of treaties), there are differences in the treaty dates across most databases. For instance, the ASEAN-China agreement is marked as having been signed in 2004 in both DESTA and WTO, whereas the WB uses 2002. These - often

Figure 1. Difference in years for each database across all three year variables. The zero mark is the harmonised year of signature in our combined data. Some databases do not appear for each year category, but that is either because we do not have the information from that data set, or because there is no difference since we prioritised information from that data set to construct the combined information.

[8] For the other treaties, we use WTO first, followed by WB and CROP's names for the remaining ones.

minor discrepancies are captured by the blue dots in Figure 1.

Although Lomé I appears as an outlier for RIKS in the date of inactivity, that is simply due to the different encoding across databases. RIKS combines all four Lomé conventions into one that covers 1975-2000, so the discrepancy is compensated for. A similar situation occurs due to different encoding convention for North American Free Trade Agreement (NAFTA), which was signed in 1992, but entered into force in 1994. DESTA and WTO use 1992 as the year of signature, but RIKS uses 1994. On the RIKS website, there is a distinction between the year of signature of the agreement, which is indeed 1992, and the year of entry into force (1994).

We add a few rules to address the discrepancies more cohesively. Namely, we ensure that all interim agreements end when their permanent version is signed. We also arbitrarily add a few end dates. For instance, the agreement Finland-USSR should end by 1991, since that is when the Soviet Union was fully dissolved. Moreover, we also make sure that agreements that involve two future EU members end as soon as one of them joins the EU. For instance, Estonia Poland (signed in 1998) is set to end in 2004, when both countries entered the EU. Similarly, agreements between a future EU member and the EU also end once the country joins the EU. However, to avoid making too many assumptions, for the RIAs where one of the two countries is not part of the EU, we do not add an end date in case it is missing. That is the case for Latvia Ukraine (signed 1995) which we leave running for the entire period even though Latvia joined the EU in 2004.

There are a few rare occurrences for which the year of entry into force precedes the year of signature of an agreement. That occurs in both the WB (Economic Cooperation Organization: signed in 2003, entered into force in 1992) and in DESTA (El Salvador Panama: signed in 1986, entered into force in 1974). When the treaty also appears in the other databases, we follow the more logical years of the other databases.

In some cases, we reconcile the differences using additional sources. Notably, the Melanesian Spearhead Group (MSG) has a different date of signature across nearly all databases: DESTA, CROP and WTO mark it as 1993, RIKS as 1986, and the WB has no date. Looking at the organisation's official website[9], we see that 1986 corresponds to an informal meeting held between the future member countries that established the MSG. Accordingly, 1993 coincides with the year of entry into force.

It is worth mentioning that the depth of RIAs is an additional piece of information we could have chosen to harmonise across the databases. Out of the databases we use, both DESTA and CROP include information on the types of provisions included. The difference in methodologies and results were too large so we did not harmonise this data.

Moreover, in earlier versions of this database, we also added the databases of Kohl et al. (2016) and Hofmann et al. (2017) to the merge. We wanted to also combine the data on the content of RIAs, as provided by these two databases, to complement DESTA and CROP. However, given the large differences in methodologies followed by the authors to create the respective databases, this merging became meaningless. For instance, we computed the correlations between all the content-related variables of DESTA and Kohl et al. (2016) for the 283 treaties that appear in both. The highest correlation (0.71) was found between DESTA's measure of depth based on a latent trait analysis and Kohl et al. (2016)'s count of provisions (regardless of their legal enforceability status). Overall, the mean of the correlations between all variables of both data sets was 0.3610. We leave the comparison between the content variables constructed by these four databases for future work.

3.1.2. Creation of Additional Variables

Parent

In many instances, RIAs can be linked via founding and amending treaties. When we identify them, we create a common "Parent" key to help recognise them. The naming convention is to use the current name of the organisation. For instance, the European Coal and Steel Community and the treaty of Maastricht both have the European Union (EU) as the parent. Table 1 summarises

[9] Retrieved from https://msgsec.info/about-msg/, last accessed on 17/10/2022.

[10] The full distribution of the correlations was rather low: the first quartile was 0.28, the median was 0.34, and the third quartile was 0.44.

the number of treaties per database, according to unique RIAs and parents. In our data, the five largest parents are the Association of South East Asian Nations (ASEAN), with 16 treaties, the African Union with 14 treaties, and the Eurasian Economic Union and Andean Community with 10 treaties.

Variable	Combined	DESTA	RIKS	CROP	WTO	WB
Total Treaties	1149	850	167	273	511	330
Total Parents	789	699	118	83	466	322

Table 2. Number of treaties and parent treaties per database. The numbers are computed post-data cleaning and merging, so there may be slight discrepancies with the original files. Note also that since there is overlap between the databases, the total number of treaties and parents does not equal the sum of the values per database.

Type: Bilateral vs plurilateral

The first variable we create distinguishes between bilateral and plurilateral treaties. We follow a simple rule: if a treaty involves two countries, it is marked as bilateral, and in all other cases, it is marked as plurilateral. Note that in most databases, such as DESTA, an RIA between a RO and a third country would be marked as bilateral, while we considered it plurilateral. Treaties between regional organisations are also categorised as plurilateral.

In total, our data has 569 bilateral RIAs and 581 plurilateral RIAs.

Type2: Trade agreement vs Regional organisation

The second variable distinguishes trade agreements from regional organisations. Both are agreements between two or more parties aimed at reducing or eliminating barriers to trade (in goods or services). The difference is that ROs also set up an institutional structure that monitors, plans or implements the treaty (either in a supranational or intergovernmental way). To make the classification, we use the fact that all treaties that appear in RIKS and CROP are ROs. We refine this variable by looking at the parent information: if at least one of the observations in the same parent group is marked as being a RO, then all other treaties within that group are also marked as ROs. Out of 1,149 observations, 358 are marked as regional organisations.

Region

The way in which the regional classification of RIAs is determined differs significantly from what is usually done. Most databases assign each country to a continent. Any RIA whose members are from different continents are then labelled as intercontinental. However, given how vague the definition of a continent is, this can lead to an overidentification of the number of intercontinental treaties. E.g. using the World Bank's definition of regions, the treaty between the EU and Malta would be considered interregional (intercontinental).

To assign the treaties to a continent, we follow a slightly more complex set of rules that better align with our expected definition of regional vs intercontinental treaties (Figure 2). To start, we follow the World Banks' devising of the world into seven regions: East Asia and the Pacific (EAP), Europe and Central Asia (ECA), Latin America and the Caribbean (LAC), Middle East and North Africa (MENA), North America (NA), South Asia (SA), and Sub-Saharan Africa (SSA).

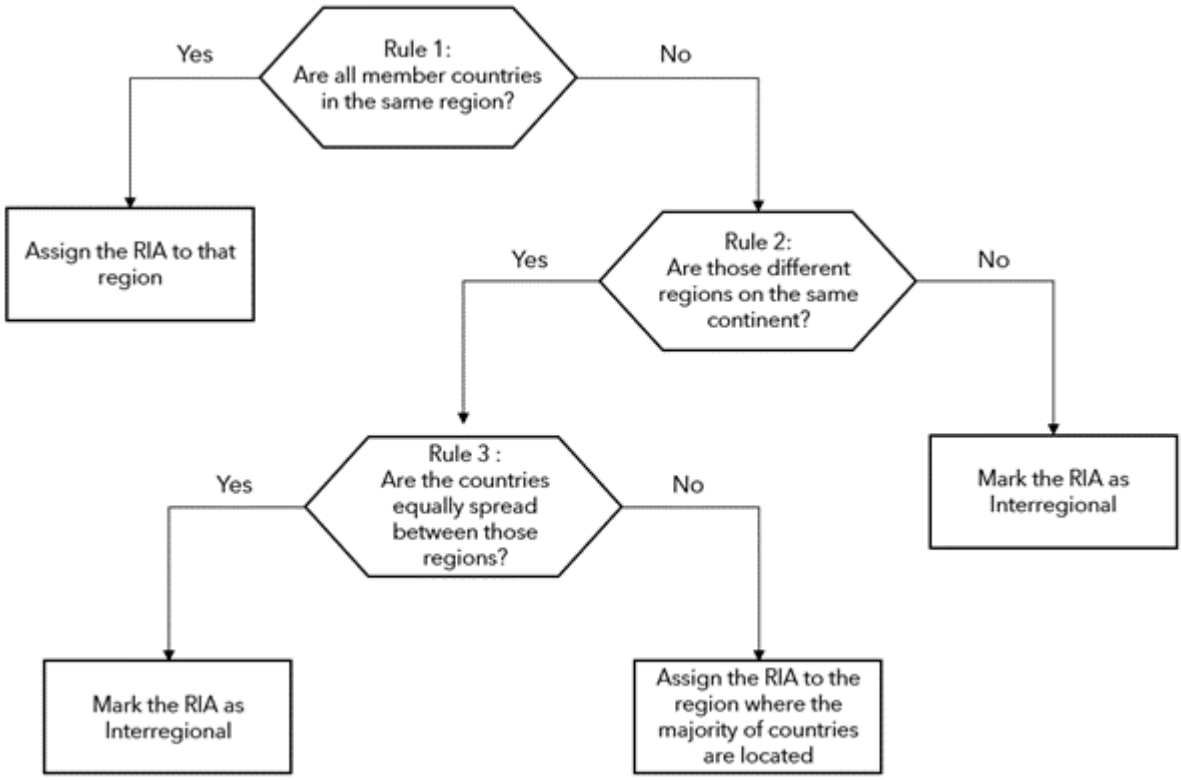

Figure 2. Decision tree to assign regions to the RIAs.

Based on rule 1, we assign a treaty such as Afghanistan India to South Asia since both countries belong to that region. Using rule 2, we can immediately assign RIAs such as Canada-Korea to the Interregional category as the former country is located in North America and the latter in East Asia and the Pacific. Rules 2 and 3 allow us to solve the puzzle illustrated by the European Union. Since Malta is located in MENA and the remaining EU countries are in ECA, we look at rule 2. Most countries in MENA border the Mediterranean Sea, as do the Southern EU countries. The two regions are also connected by land, so we consider them to be on the same continent. Via rule 3, we implement a majority rule and assign the EU – and countries in a similar scenario – to ECA.

The final division of RIAs based on both regions and year of signature can be found in Figure 3. We see that TAs signed in the 90s were predominantly signed in Europe and Central Asia. Regional organisations were regularly set up across all decades, but Interregional trade agreements mainly became popular post 2000. Sub-Saharan African countries mainly engage in regional organisations but not as much in trade agreements. LAC countries actively signed both types of RIAs between 1980 and 2009. They were a bit more active on the global trading scene pre-1980 and post-2009 via regional organisations than trade agreements. We can also see that the golden age of RIAs took place between 1990 and 2009.

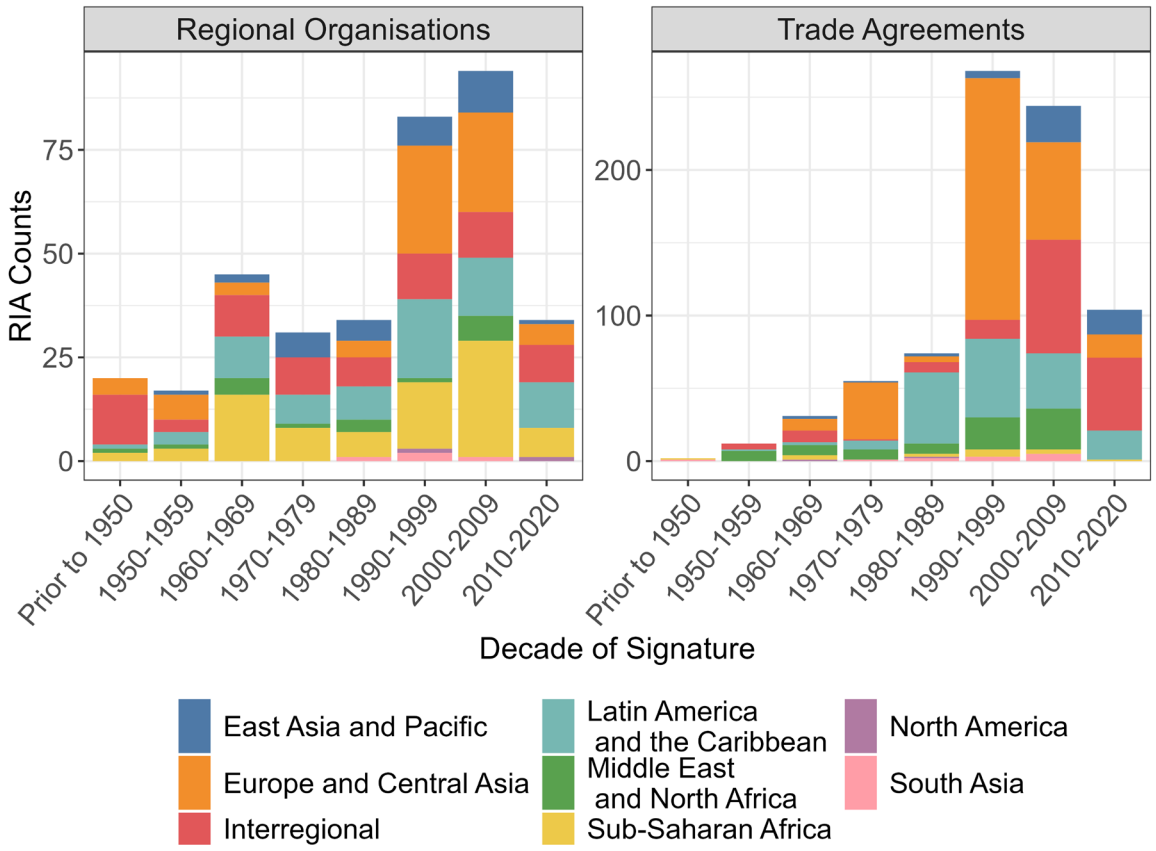

Figure 3. RIA counts across regions and years of signature.

3.2 Membership and Dyads Data

Building on our combined data set, we also construct a membership data set and a dyads data set. The former lists all member states to an RIA in a given year. The second reproduces the same information but in a country pair-year format. When available, we use the year of entry into force. Both files include a unique identifier key that allows the data to be combined with the more complete treaty information.

These additional files thus offer information on the year of accession and withdrawal of countries to RIAs. It is important to note that to construct those two additional files, we do not use information from CROP since nearly all the treaties included about ROs are captured via RIKS.

3.2.1 Membership Data

In the membership file, we collect information on country participation in the RIAs. We do so for each year a treaty is in effect. As such, we capture the entry and exit to a treaty (accession/withdrawal) as well as the year of inactivity/end of an RIA if applicable. To collect the data, we primarily rely on RIKS and DESTA as they provide data in a format easily tailored to our needs. We manually add member countries for the remaining databases based on the treaty information.

The final file contains membership information on 235 countries and territories, spanning 1910-2020.

3.2.2 Dyads Data

To construct the dyads data, we start from our membership file and transform it in a country-country format. For instance, if a treaty has ten member countries in the membership file, it will contain 45 country pairs (n*(n-1)/2 = 10*9/2).

To prevent over-inflating the data, we do not duplicate internal RO links for RIAs between ROs or between ROs and third countries. For example, the treaty between the European Free Trade Association (EFTA) and Georgia will have four links (one for each member country's connection with Georgia) in our dyads file. It does not repeat the internal EFTA links already accounted for under the main EFTA treaty. Without this correction, there would have been ten links for this treaty.

To implement this correction, we check DESTA as it contains information on the type of agreement according to its membership: "plurilateral & third country" and "region-region." For those treaties not included in DESTA, we also consider the type and RTA composition provided by the WTO. For the remaining treaties that do not appear in either, we check them manually.

4. Value-Added of the Combined Data

In this section, we first review the scope of our combined data set compared to the original databases. Then, we run some analyses comparing all data sets to study the value-added of the data.

4.1 Scope of the Combined Data

First, we focus on the level of overlap between the databases. In this case, we define overlap as the number of treaties that appear in more than one database. The results show that nearly 45% of all treaties only appear in one database, with over 25% coming solely from DESTA. Without DESTA, over 300 treaties are not included in analyses. The WB only contributes 7 treaties that do not appear in any other database. In total, there are only 14 treaties that can be found in all five databases, and they are all regional organisations. The highest overlap (20.1%) occurs between the databases focusing on trade agreements (DESTA, WTO, and WB). Between pairs of databases, the combination DESTA-WTO has the highest overlap (15.2%).

Interestingly, the WTO contains 32 treaties that do not appear in any other database. Upon closer examination, these treaties can be divided into two main categories. Most of those are interim agreements signed between the EU and third countries, ending after a few years as they become superseded by their definitive versions. The second category contains treaties between Central American countries and third countries. The WTO creates a treaty for each Central American country, whereas the other databases tend only to capture the regional-third country treaty. The remaining treaties (5 out of 32) that appear solely in WTO do not appear to follow any clear pattern.

In the upper panel of Figure 4, we can analyse the coverage by decade. Over the first eight decades of the combined data, DESTA provides the most complete source of data. In the last category, the WTO takes over because it was updated more recently than DESTA. The coverage of the WTO varies over the decades. Between the signature of the earliest treaties and 1989, very few treaties were notified to the WTO, except 1970-1979, where 50% of the treaties signed were notified to the WTO. The WB only has good coverage for the years 1999-2009. From this graph, we can also see that nearly all of the regional organisations formed before 1950 are only covered by RIKS and CROP.

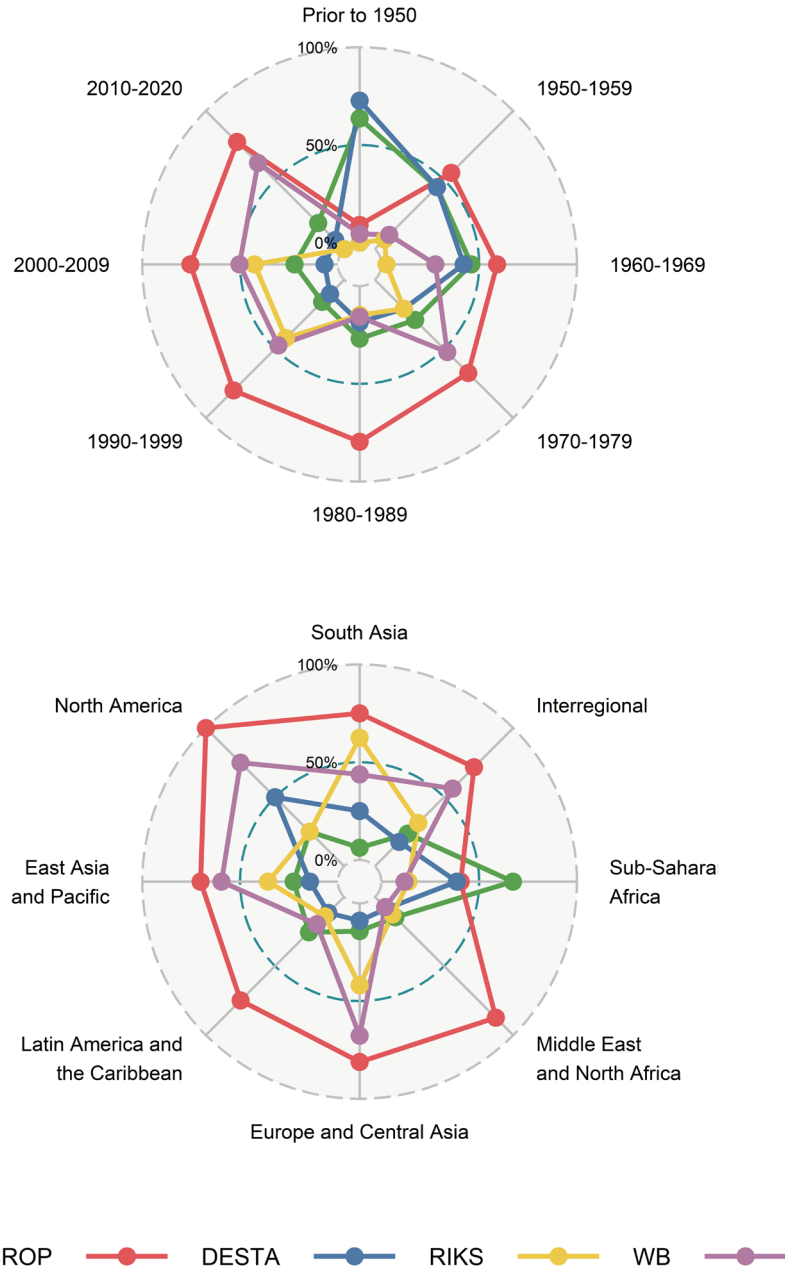

Figure 4. Combined database coverage by original database, broken down by decade of signature (upper panel) and by treaty region (lower panel).

Turning our attention to the coverage by regions (lower panel of Figure 4), we see that DESTA contains over 50% of the treaties across all regions except for Sub-Saharan Africa. As to the WTO, we see that in half of the regions, less than 50% of all treaties are notified to the WTO. The regions with the most notifications are North America, Europe, and Central Asia. Only for South Asian treaties does the WB provide significantly more information than the WTO.

4.2 Comparative Analyses

To assess the need for such a combined data set, we run two types of analyses. The first one consists in a basic network analysis and the second one in a gravity model. To conduct these analyses, we mainly work with the dyads we obtained from the membership information. From this point onwards, we exclude CROP from the comparisons, since the database does not provide membership data.

4.2.1 Network Analysis

The dyads file we construct allows us to easily use tools from network analysis. The first measure we look at is the density of our data. In other words, we compute the proportion of countries that are linked by at least one RIA in a given year compared to the number of possible links that could be present between all countries and territories. The density provide a rough estimate of the level of global integration.

The results are displayed in Figure 5. In both panels we see that the combined data provides a larger coverage of countries. In the upper panel, the density calculated using RIKS is very high due to the inclusion of organisations with a non-economic focus, and it drives the density of the combined data. We exclude those in the lower panel[11] and compute the density again. When we restrict the analysis to organisations with an economic focus, we observe a steady rise in density across all databases. However, in the 90s, we notice that the increase in density obtained with the WTO data is less pronounced. Specifically, the densities for the other databases increase by 5-7% in that decade, whereas for the WTO, it only increases by over 2%. From the 2000s onward, both RIKS and the WB's densities slowly stagnate, whereas for DESTA and the WTO's (and thus the combined data), it keeps increasing.

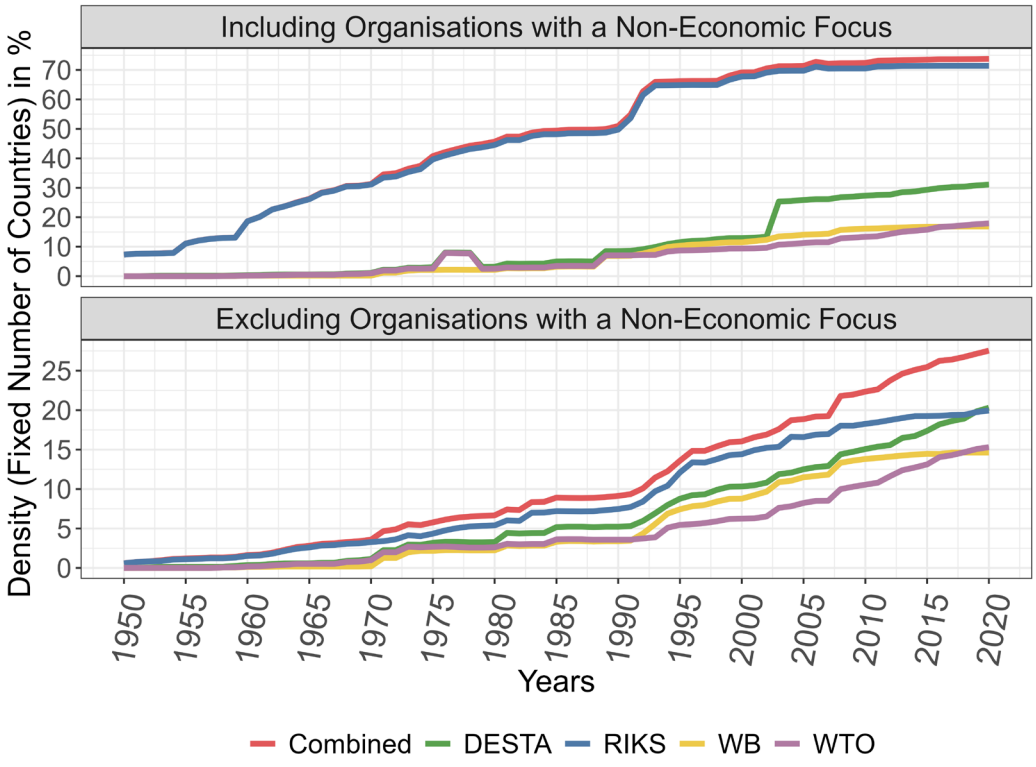

Figure 5. Evolution of the density based on the source of the data used.

[11] The full list of treaties we removed can be found in 6.2 Appendix B – List of Agreements and Organisations without an Economic Focus. Without those, the combined data contains 1,071 treaties instead of 1,149.

The second measure we consider is the degree centrality of countries. Degree centrality counts the number of treaties linking a country to other countries in a given year. If a country is part of the same RIA with two other countries, its degree centrality will be two. Alternatively, if a country has a treaties in force with two different countries, its degree centrality will also be two. Since the scope of agreements covered largely differs between databases, we compare the top five countries based on their degree centrality from each database. The results are displayed in Figure 6.

In the combined database, all five leading countries are located in ECA. The Netherlands comes out as being the most central country. However, no other database has it in its top five. The United Kingdom also only appears in this top five. France appears in all databases, but not always in the same spot. Germany and Italy appear in some of the databases. Egypt, Belgium, Tunisia, and Morocco appear in two or three databases' rankings, but not in our database. Algeria and Luxembourg only come in as fifth in RIKS and WTO, respectively. Overall, the average degree centrality in 2020 is highest in the combined database (66). This is logical as our data includes more treaties. RIKS and DESTA come in second in terms of average degree with 50. For the WTO, the average only reaches 40, and for the WB, it is only 36. These differing results highlight the sensitivity of this type of analysis to the data selected.[12]

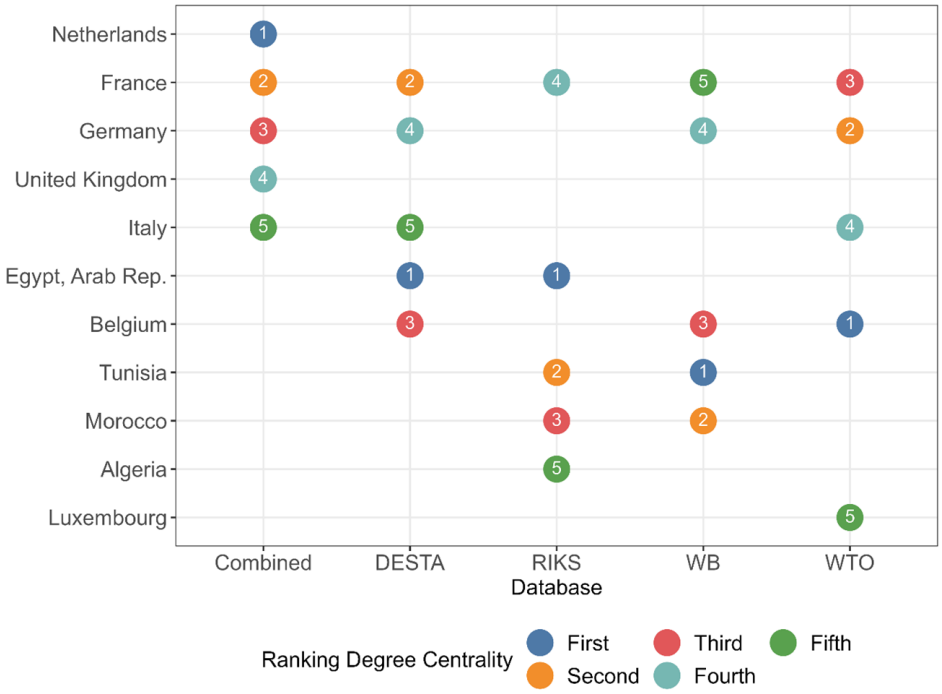

Figure 6. Ranking of the countries' degree centrality obtained with each of the database in 2020. Note that a country can be ranked lower than another in a different database since the absolute values differ.

To go beyond the top five countries, we also look at the distribution of the degree centrality across each database (Figure 7). Most of the databases display a right skew meaning that there are more countries with a lower degree than countries with a very high degree centrality. This is in line with network studies. This pattern is somewhat less pronounced for RIKS and the WB. DESTA and our combined database both display three clusters of values, while the other databases tend to only have two. The highest cluster in DESTA, WTO, and our data all indicate the presence of a few hubs that actively engage in RIAs.

[12] Note that the degree centralities computed with the combined database and the original ones are all significantly different.

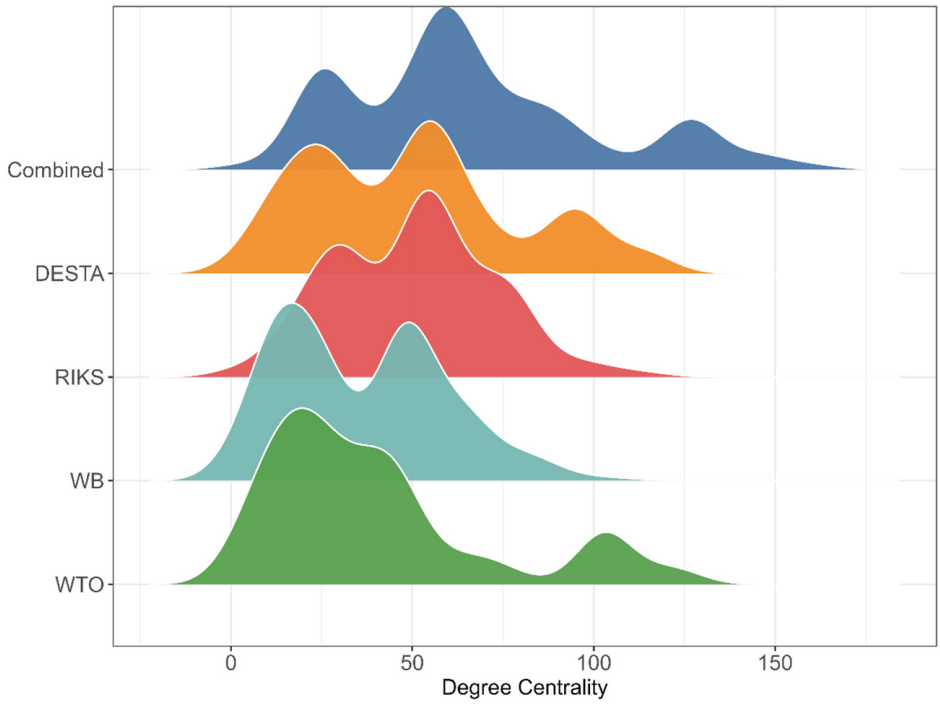

Figure 7. Comparison of the density plots of the degree distributions obtained with each database for 2020.

4.2.2 Gravity Model

To further illustrate the need for an exhaustive database, we use them to estimate a simple gravity model, the workhorse of international trade studies (Yotov, 2022). Specifically, we estimate the impact of the treaties on the bilateral flow of trade using the different datasets. Our goal is not to provide evidence of which dataset is superior but to illustrate the impact of the change in coverage on these estimations.

Once again, we only consider RIAs with an economic focus. We estimate the gravity model in accordance with the latest preferred specification: i.e. including importer-year, exporter-year and importer-exporter fixed effects and using a Pseudo Poisson Maximum Likelihood estimator. In each regression, we assume that if the database does not mention the existence of a treaty, there is no treaty active.

The results of the gravity model show that there are large discrepancies in the estimated partial equilibrium effects (Table 3). The largest effect is obtained under RIKS, where a treaty is expected to raise the flow of trade by more than (= 151%, followed closely by the combined dataset (143%). In contrast, when using DESTA and the WTO the effect of a treaty shrinks considerably: 22% or 30%. The likely reason for these differences is straightforward. The difference between countries with and without a treaty is used to estimate their impact, so the choice of which treaties to include is not without effect. Working with an exhaustive database will likely limit this coverage discrepancy's impact.

VARIABLES	(1) Combined	(2) DESTA	(3) WTO	(4) WB	(5) RIKS	(6) CROP
RIA	0.890***	0.265***	0.201***	0.383***	0.921***	0.557***
	(0.0150)	(0.0125)	(0.0119)	(0.0118)	(0.0159)	(0.0190)
Constant	27.52***	27.60***	27.60***	27.60***	27.52***	27.57***
	(0.00205)	(0.00157)	(0.00152)	(0.00150)	(0.00209)	(0.00202)
Observations	874,745	874,745	874,745	874,745	874,745	874,745

Table 3. Estimation results of gravity models with import-year, export-year, and panel fixed effects, estimated with Poisson pseudo maximum likelihood. The robust standard errors are in parentheses.
Significance thresholds: *** p<0.01, ** p<0.05, * p<0.1.

5. Conclusion

In this working paper, we have demonstrated the benefits of combining different data sets on RIAs. In this manner, we can leverage additional information with higher accuracy while avoiding the limitations associated with specific databases. The three files that we constructed, the combined data, membership data, and dyads data, aim to capture all aspects of both regional organisations and trade agreements, including characteristics at the RIA level and information at the country level. The combined data we create includes information that is not fully contained in any single database we build on. Moreover, the coding system we use makes it easy to navigate across the treaties in all three files and also allows to search information from each of the original databases. Furthermore, we have highlighted the level of bilateral links between countries that are lost, as well as the regions and decades that are most affected by missing information when we consider these databases separately. We have also shown how sensitive analyses are to the selection of data, thus requiring actions to prevent biases.

Despite the progress made in this paper, there are still a few avenues left for further research. First, more databases could be considered. Even though we mentioned a few additional ones, this analysis could be repeated including those other databases. In this manner, additional information on the characteristics of RIAs could be captured by the combined data. Second, the manual part of the merging procedure could be more systematic. Unfortunately, due to the name- and date-specific discrepancies mentioned throughout the paper, it is difficult to automate this step. Nevertheless, accounting for the observations made here, a more standardised approach could be developed. Finally, building on the more exhaustive list of RIAs provided here, one could attempt to code the depth of all RIAs using a unique methodology to obtain the richest source of information on the global level of integration.

References

Auguie B (2017). _gridExtra: Miscellaneous Functions for "Grid" Graphics_. R package version 2.3, <https://CRAN.R-project.org/package=gridExtra>.
Baier, S. L., Bergstrand, J. H., & Mariutto, R. (2014). Economic determinants of free trade agreements revisited: Distinguishing sources of interdependence. *Review of International Economics*, *22*(1), 31–58.

Baier, S. L., Yotov, Y. V., & Zylkin, T. (2019). On the widely differing effects of free trade agreements: Lessons from twenty years of trade integration. *Journal of International Economics*, *116*, 206–226.

Bion R (2023). _ggradar: Create radar charts using ggplot2_. R package version 0.2.

Comtois D (2022). _summarytools: Tools to Quickly and Neatly Summarize Data_. R package version 1.0.1, <https://CRAN.R-project.org/package=summarytools>.

Csardi G, Nepusz T: The igraph software package for complex network research, InterJournal, Complex Systems 1695. 2006. https://igraph.org

Dowle M, Srinivasan A (2022). _data.table: Extension of `data.frame`_. R package version 1.14.6, <https://CRAN.R-project.org/package=data.table>.

Dür, A., Baccini, L., & Elsig, M. (2014). The design of international trade agreements: Introducing a new dataset. *The Review of International Organizations*, *9*(3), 353–375.

Egger, P., & Larch, M. (2008). Interdependent preferential trade agreement memberships: An empirical analysis. *Journal of International Economics*, *76*(2), 384–399.

Harrell Jr F (2022). _Hmisc: Harrell Miscellaneous_. R package version 4.7-2, <https://CRAN.R-project.org/package=Hmisc>.

Hofmann, C., Osnago, A., & Ruta, M. (2017). Horizontal depth: A new database on the content of preferential trade agreements. *World Bank Policy Research Working Paper*, *7981*.

Horn, H., Mavroidis, P. C., & Sapir, A. (2010). Beyond the WTO? An anatomy of EU and US preferential trade agreements. *The World Economy*, *33*(11), 1565–1588.

Jetschke, A., Münch, S., Cardozo-Silva, A. R., & Theiner, P. (2021). Patterns of (Dis) similarity in the Design of Regional Organizations: The Regional Organizations Similarity Index (ROSI). *International Studies Perspectives*, *22*(2), 181–200.

Kassambara A (2022). _ggpubr: 'ggplot2' Based Publication Ready Plots_. R package version 0.5.0, <https://CRAN.R-project.org/package=ggpubr>.
Kohl, T., Brakman, S., & Garretsen, H. (2016). Do trade agreements stimulate international trade differently? Evidence from 296 trade agreements. *The World Economy*, *39*(1), 97–131.

Limão, N. (2016). Preferential trade agreements. In *Handbook of commercial policy* (Vol. 1, pp. 279–367). Elsevier.

Mario Larch's Regional Trade Agreements Database from Egger and Larch. (2008). [Data set]. https://www.ewf.uni-bayreuth.de/en/research/RTA-data/index.html

Müller K (2020). _here: A Simpler Way to Find Your Files_. R package version 1.0.1, <https://CRAN.R-project.org/package=here>.

NSF-Kellogg Institute. (n.d.). *Database on Economic Integration Agreements* [Data set]. NSF-Kellogg Institute. Retrieved December 12, 2022, from https://sites.nd.edu/jeffrey-bergstrand/database-on-economic-integration-agreements/

Orefice, G., & Rocha, N. (2014). Deep integration and production networks: An empirical analysis. *World Economy*, *37*(1), 106–136. https://doi.org/10.1111/twec.12076

Schauberger P, Walker A (2022). _openxlsx: Read, Write and Edit xlsx Files_. R package version 4.2.5.1, <https://CRAN.R-project.org/package=openxlsx>.s

Slowikowski K (2022). _ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'_. R package version 0.9.2, <https://CRAN.R-project.org/package=ggrepel>.

UNU-CRIS. (n.d.). *Regional Integration Knowledge System* [Data set]. United Nations University Institute on Comparative Regional Integration Studies. Retrieved December 12, 2022, from https://riks.cris.unu.edu/

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). "Welcome to the tidyverse." _Journal of Open Source Software_, *4*(43), 1686. doi:10.21105/joss.01686 <https://doi.org/10.21105/joss.01686>.

Wickham H, Miller E, Smith D (2022). _haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files_. R package version 2.5.1, <https://CRAN.R-project.org/package=haven>.

Wilke C (2022). _ggridges: Ridgeline Plots in 'ggplot2'_. R package version 0.5.4, <https://CRAN.R-project.org/package=ggridges>.

World Bank. (n.d.). *Global Preferential Trade Agreements Database* [Data set]. World Bank. Retrieved December 12, 2022, from https://wits.worldbank.org/gptad/library.aspx

World Trade Organisation. (n.d.). *WTO - Regional trade agreements Database* [Data set]. World Trade Organisation. Retrieved December 12, 2022, from https://rtais.wto.org/UI/PublicMaintainRTAHome.aspx

Yotov, Y. V. (2022). *Gravity at Sixty: The Workhorse Model of Trade.* cesifo working papers n. 9584. https://www.econstor.eu/bitstream/10419/252101/1/cesifo1_wp9584.pdf

6. Appendices

6.1 Appendix A – Special Cases

For certain treaties, we encountered specific issues that required a few judgment calls. This section lists those exceptions:

- The treaty Central Europe Free Trade Agreement (CEFTA)-Croatia (signed in 2002 with ratification in 2003) from the WB corresponds to the accession of Croatia to CEFTA, so we exclude the treaty from the WB database when merging with other databases.
- The treaty Chine China, from the WB, appears only once in all other databases, so even though the years of signature are rather different, we still merge the treaties
- We match the treaty between El Salvador and Panama even though DESTA says it was signed in 1986, and WB claims it was signed in 1970, as they seem to be linked with each other at the very least.
- We exclude the treaty between the United States and Albania (signed in 1995 with ratification in 1998) from the WB. In the WB, it appears as a bilateral free trade agreement, but upon closer look, it is a bilateral investment treaty, which is a type of treaty that we do not consider here.
- The treaty between India and Mongolia, in the WB database, has no signature date. We found an agreement text that seems to match this treaty with a signature date of 1996.
- The amending treaty of the Southern African Development Community (SADC) from the WB has no date nor any document. So we match it with one of the CROP treaties that revises the SADC.
- Even though WTO and DESTA are the two most complete databases we build on, we identify a few key differences between them. First, both differ in how they mark accession and withdrawal treaties. Second, we note the absence of a few treaties in DESTA despite their notification to the WTO. Similarly, DESTA includes a variable that indicates whether an treaty also appears in the WTO[13], which appears to not always be accurate. In particular, there are 26 treaties, including the DESTA treaties between Cuba and Mexico and between Australia and Indonesia, which are marked as not appearing in the WTO in any form, even though the WTO does include them. This is likely due to the late notification of the large majority of the agreements to the WTO. Indeed, most of the notifications took place post-2019, so they have perhaps not yet been included in DESTA.
- DESTA has two treaties involving the EU that appear to be duplicated (EC Vietnam - 2019, and EC Singapore - 2018), since only one version appears in the WTO. Therefore, we only keep one of the two treaties and match them when possible.
- The WB appears to have repeated treaties. For instance, Afghanistan - India (2003) appears three times. In two of those three cases, the treaties are identical across all characteristics provided by the WB database, and the third one differs in a couple of ways. When we encounter this and can say with a certain degree of confidence that the treaty was duplicated by comparing this treaty with its existence in other databases, we only keep one of the treaties.
- The WB database has a lot of missing dates that sometimes makes the matching procedure more complex. When there is a date of entry into force, we compare it to the other databases to check for possible matches.

[13] This variable is rather broad and captures treaties that appear in the WTO either as a RTA in force, an inactive agreement, an accession treaty or an early announcements/under negotiation treaty.

6.2 Appendix B – List of Agreements and Organisations without an Economic Focus

African Union
African, Caribbean and Pacific Group of States
Arab League
Belarus Russia (Union State)
Casablanca Group
Cotonou Agreement
Council of the Entente
European Atomic Energy Community
General Agreement on Tariffs and Trade
Global System of Trade Preferences (GSTP)
Group of 77
GUAM/GUUAM Organization for Democracy and Economic Development
Indian Ocean Rim Association
Intergovernmental Group of Twenty-Four on International Monetary Affairs and Development
International Labour Organization
International Monetary Fund
League of Nations
Lome I
Lome II
Lome III
Lome IV
Monrovia Group
Non-Aligned Movement
North Atlantic Treaty Organization
Opec Fund for the International Development
Organization for Economic Cooperation and Development
Organization for European Economic Co-operation
Organization for Security and Cooperation in Europe
Organization of African Unity
Organization of American States
Organization of Islamic Conference
Protocol to the Treaty on the establishment of the Central American Parliament
Union State of Russia and Belarus
United Nations
United Nations Conference on Trade and Development
United Nations Department of Economic and Social Affairs
United Nations Development Programme
United Nations Economic and Social Commission for Asia and the Pacific
United Nations Economic Commission for Latin America and the Caribbean
Western European Union
World Bank Group
World Trade Organization

Towards an Exhaustive Regional Integration Agreements Database

UNU-CRIS Working Paper #2 2023

Published by: United Nations University Institute on Comparative Regional Integration Studies

Cover image: John Lockwood/Unsplash