


ORIGINAL ARTICLE

# Introducing the Electronic Database of Investment Treaties (EDIT): The Genesis of a New Database and Its Use

Wolfgang Alschner<sup>1</sup>, Manfred Elsig<sup>2\*</sup>, and Rodrigo Polanco<sup>3</sup> 

<sup>1</sup>University of Ottawa, Canada, <sup>2</sup>University of Bern, Switzerland and <sup>3</sup>University of Bern, Switzerland

\*Corresponding author: Email: [Manfred.Elsig@wti.org](mailto:Manfred.Elsig@wti.org)

(Received 27 February 2020; revised 25 May 2020; accepted 6 June 2020; first published online 13 August 2020)

## Abstract

This article introduces a novel database on investment treaties called the Electronic Database of Investment Treaties (EDIT). We describe the genesis of the database and what makes EDIT the most comprehensive and systematic database to date. What stands out besides the coverage is that treaties are all provided in one single language (English) and in one single format that is machine-readable. In the second part of the article, we provide selected illustrations on how the data can be used to address research questions in international law, international political economy, and international relations by applying text-as-data methods and by extracting and visualizing data based on EDIT.

**Keyword:** Investment; Treaties; Database; International Relations; International Law; International Investment Agreements

## 1. Introduction

The international investment agreement (IIA) universe has become a fixation for policy-makers, civil society, and scholars across disciplines. The global proliferation of bilateral investment treaties (BITs) and preferential free trade agreements (PTAs) with investment components has created a web of more than 3500 treaties, most of which offer private investors the possibility to launch direct claims for treaty violation against sovereign states (UNCTAD, 2019). As a result, the IIA regime has become one of the most litigated and controversial areas of international law with close to 1000 disputes (UNCTAD, 2019). These disputes, at times, challenge sensitive areas of public policy-making from tobacco control to renewable energy policies and can lead to multi-billion dollar decisions (or ‘awards’ as they are known in arbitration) against host states (UNCTAD, 2015).

In consequence, citizens have protested against investment arbitration, states have denounced or renegotiated their investment treaties, and policymakers across the globe have begun to rethink the system, culminating in ongoing multilateral reforms under the auspices of the United Nations. Scholars have been equally captivated by investment agreements. Political scientists have asked why states sacrificed their sovereignty (Poulsen, 2015), economists have investigated whether these treaties yield the investment-promoting benefits they promise (Bellak, 2015; Desbordes, 2016; Egger *et al.*, 2019; Pohl, 2018; Sauvage and Sachs, 2009; Yackee, 2008, 2010), and legal scholars have studied the burgeoning investment arbitration case law that interprets and applies these agreements (see e.g. Brown, 2011).

In spite of the intense interest IIAs have sparked and the research they have motivated, important aspects of the IIA universe still remain unknown or understudied. Scholars have focused attention on readily available English language treaty texts of major capital importing and

exporting states leaving the IIA practice of entire regions, mostly in the developing world, underexplored. Furthermore, although coding efforts by international organizations and researchers have elucidated important general content variations among IIAs, treaty language particular to specific countries or of interest to particular research questions have remained outside of their purview. As a result, the collective understanding of the IIA universe is patchy, prone to misconceptions, and partially biased in favor of English-language treaty practice of dominant states.

To address these shortcomings, this article introduces the Electronic Database on Investment Treaties (EDIT). EDIT seeks to facilitate informed public discourse, evidence-based reform, and academic study by providing unprecedented access to the most extensive set of the IIA texts to date. EDIT stands out among existing investment treaty resources for five reasons: it is (1) comprehensive, (2) uniform, (3) machine-readable, (4) annotatable, and (5) freely accessible to researchers and the public.<sup>1</sup>

First, EDIT is the most comprehensive database on IIA full texts to date. Relying on more than 3474 IIAs available texts, it contains more agreement full texts than any other database.<sup>2</sup> In the process of creating EDIT, we contacted national agencies and ministries and uncovered hitherto unavailable full texts shedding new light on the treaty practice of states, e.g. in the Middle East and North Africa, which are underrepresented in other datasets. In addition, we adopt a more inclusive definition of IIAs that not only includes BITs and PTAs, but also Friendship Commerce and Navigation (FCN) treaties, which have been coined the ‘first investment agreements’ (Vandeveldel, 2017). As a result, EDIT offers the most comprehensive overview of the IIA universe yet.

Second, EDIT is uniform insofar as it contains an English version of each treaty. While EDIT includes more non-English texts than any other investment law database, we realized that Arabic, Russian, and German texts may not be accessible to all users. Therefore, we provide an English translation of all non-English treaty texts. While the treaty’s original language is available in the database and constitutes the authentic text, the standardized translation allows users to search for key provisions such as ‘fair and equitable treatment’ or ‘full protection and security’ across the entire database irrespective of the original language. We thereby hope to make the content of understudied non-English texts more widely accessible.

Third, EDIT is machine-readable. Existing databases often contain imaged texts that can be read by humans but not searched or analyzed by computers. EDIT now renders all IIAs machine-readable. We went through the painstaking process of digitizing texts through optical character recognition (OCR) and manually corrected ensuing conversion errors. We also standardized the structure of each treaty through Extendable Markup Language (XML) and Extensible Hypertext Markup Language (XHTML) that allows us to distinguish between preambles, articles, and annexes. As a result, users can now search through all agreements more efficiently. In addition, EDIT allows for the machine-processing of text-as-data to support the emerging field of computational legal research.

Fourth, EDIT can be annotated to map the content features of treaties, meaning that leveraging the XML structure of EDIT documents, users can identify and mark content features of investment agreements directly in the text. We use a transparent and automated keyword-based approach to identify a first layer of content features that are likely to be of general interest to users and researchers. In addition, we anticipate that users will validate, refine, and add to this content mapping over time through their own annotations. In other words, EDIT allows crowd-sourcing the content analysis of IIAs.<sup>3</sup> Furthermore, while the authors and the World Trade Institute, in

<sup>1</sup><https://edit.wti.org>. EDIT is a free academic resource provided by the World Trade Institute – University of Bern. It was developed during the implementation of the SNIS-funded project ‘Diffusion of International Law: A Textual Analysis of International Investment Agreements’ (2015–2017), in cooperation with its research partners, the University of Ottawa through funding from the Canada Foundation for Innovation and the World Bank.

<sup>2</sup>It must be noted that we exclude general framework agreements counted by UNCTAD as a treaty with investment provisions (TIPs), since they do not themselves contain specific investment protection or liberalizations provisions.

<sup>3</sup>A special protocol how to manage future automated and manual annotation is being developed in order to allow for a transparent and efficient process.

collaboration with its research partners,<sup>4</sup> are committed to regularly updating EDIT, we hope this crowd-sourcing approach will help build a broader community of users that will expand and maintain the database.

Fifth, EDIT is a free-of-charge academic resource primarily aimed at researchers. Several IIA databases are only available to paid subscribers, thus limiting access to investment treaty knowledge and expertise. UNCTAD's investment policy hub stands out as an invaluable and free resource for developing countries and the wider public. Yet, its text versions so far do not always match the quality of commercial providers and it is difficult for researchers to search and access its text in bulk. EDIT is not meant to compete with these databases, rather, it seeks to complement them by broadening access to IIAs and by facilitating innovative text-as-data research.

The primary purpose of this article is to introduce the EDIT database. In addition, we also offer a range of new descriptive insights about the investment treaty universe to showcase EDIT's utility for legal scholars, political scientists, and economists. A better understanding of the IIA universe is crucial in all three disciplines. While we anticipate that legal researchers will use EDIT to search for key terms, its unprecedented scope also provides new avenues for empirical scholarship, e.g. by studying the evolution of key clauses such as 'most favored nation treatment' provisions or by investigating the interaction between treaties and investment arbitration. Moreover, its machine-readable texts can fuel new computational legal analyses of investment treaty content (Alschner *et al.*, 2017). Political scientists, in turn, can use EDIT to investigate the determinants of investment treaty design by exploring, for example, why some content features proliferate while others decline. Finally, to economists EDIT offers multiple ways to model the design of IIAs in order to assess its effects, e.g. on investment flows and other economic factors.

This article is structured as follows. First, it outlines in more detail the added value that EDIT can provide. Second, it introduces specific EDIT features. Third, it showcases interdisciplinary applications based on EDIT to address diverse questions across multiple levels of analysis ranging from the global to the treaty provision level.

## 2. Why Another IIA Database?

EDIT goes beyond existing IIA databases. UNCTAD, for example, administers a free and extensive database and full text repository of IIAs. By May 2020, it included information on 3712 IIAs. However, the full texts of 486 IIAs are missing (around 13% of the total), leaving 3226 available texts. The large majority of the available texts are provided in English and the files are offered in pdf format, partly as scanned images and partly as machine-readable texts.<sup>5</sup> Other existing public international databases, such as the United Nations Treaty Collection<sup>6</sup> and ICSID's 'Investment Laws of the World' (the latter only in printed loose-leaf), also provide selected full texts of investment treaties, without aspiring to include all existing IIAs.

In addition, several commercial databases have been created, which are only available to paid subscribers. Kluwer Arbitration 'BITS' features around 1400 treaties supplied by the Penn State Institute of Arbitration.<sup>7</sup> Similarly, Oxford University Press (OUP) 'Investment Claims' includes around 1850 BITS.<sup>8</sup> However, both are limited to agreements in English-language. More recently, another commercial database (Jus Mundi) includes around 2500 IIAs, in several languages.<sup>9</sup>

<sup>4</sup><https://edit.wti.org/treaty/team>.

<sup>5</sup>UNCTAD, International Investment Agreements Navigator, <https://investmentpolicy.unctad.org/international-investment-agreements>.

<sup>6</sup>United Nations Treaty Collection, <https://treaties.un.org/>.

<sup>7</sup>Kluwer Arbitration Blog, Bilateral Investment Treaties (BITS), <http://www.kluwarbitration.com/bits>.

<sup>8</sup>Investment Claims, Oxford University Press, <https://oxia.oup.com/home/ic>.

<sup>9</sup>Jus Mundi, <https://jusmundi.com/en/>.

Some of the above databases also rely on investment obligations in PTAs, but lack a comprehensive coverage.

A third group of datasets have focused on coding specific provisions in investment treaties and trade agreements instead of providing the full text. In BITSEL ('BIT Selection Index'), eleven key elements have been coded manually relying on more than 1,500 BITs and more than 100 PTAs with investment chapters (Chaisse and Bellak, 2015). In DESTA ('Design of Trade Agreements'), the coding includes seven of the most relevant provisions with respect to non-discrimination, market access, and investor-state dispute settlement (ISDS), from approximately 1000 PTAs signed between 1947 and 2018 (Dür *et al.*, 2014). PluriCourts, the Centre for the Study of the Legitimate Roles of the Judiciary in the Global Order at the University of Oslo, has been coding investment treaties invoked in investor-State arbitration cases as a part of a larger project on an investment treaty arbitration database (PITAD – PluriCourts Investment Treaty Database).<sup>10</sup> Finally, the IIA Content Mapping Project, a collaborative initiative between UNCTAD and several universities worldwide to code the features of IIAs, has categorized more than 120 provisions for around 2500 treaties found in UNCTAD's database.<sup>11</sup>

Together these databases are extremely helpful, but they also suffer from several limitations. First, they lack texts of agreements currently in force, particularly older agreements and those signed by developing countries. In addition, these databases tend to focus on English treaty texts. Even though the most common language of IIAs is English (used in around 64% of the IIAs), more than 633 treaties are only available in another language (Polanco Lazo *et al.*, 2018). Several existing databases include IIAs exclusively in English, and the ones that include treaties in another language usually do not provide an English translation of the official agreement. Not accounting for these biases may affect the validity of research findings based on existing databases.

Second, some databases suffer from lack of easy access or may not be user-friendly. Commercial databases allow sophisticated text searches but are only available to subscribers. Furthermore, they typically discourage, prevent, or prohibit bulk download of texts or web scraping, which curtails the ability of researchers to run text analysis software across agreements. Publicly available databases are more accessible, but they typically contain inferior search functions and poorer data formats. Texts are often stored as image or photographic presentations. This limits the full exploitation of the data for research purposes as texts need to be converted into a machine-encoded text formats (e.g. using Optical Character Recognition (OCR)) to carry out textual analyses using computational approaches.

Third, datasets of manually coded content features are useful, but have significant shortcomings, too. They operate at a high level of abstraction and might not always detect subtle and important differences. For example, knowing that a treaty contains a clause on 'expropriation' is helpful, but even such 'standard' provisions in investment treaties can vary substantially. Indeed, even the most fine-grained manual coding efforts can never be fully satisfactory, as country-specific practices, new arbitral interpretations, and evolving research questions constantly expand the list of potentially meaningful treaty variation. Moreover, manual content mapping tends to focus on key words found in the main body texts of treaties, but does not generally account for the interaction of obligations with exceptions or additional annexes or protocols or schedules that may alter or qualify the content of obligations.

In short, there is need for a database that addresses these shortcomings and limitations. EDIT fills this gap by providing the most comprehensive and uniform database of IIAs in a machine-readable format that allows for efficient search, text-as-data analysis, and content feature annotation.

<sup>10</sup>PluriCourts, <https://pitad.org/index#welcome>.

<sup>11</sup>UNCTAD, Mapping of IIA Content, <https://investmentpolicy.unctad.org/international-investment-agreements/ii-mapping>.

### 3. Introducing EDIT

As mentioned, five features set EDIT apart from other IIA databases. EDIT is (1) comprehensive, (2) uniform, (3) standardized and machine-readable, (4) annotatable, and (5) free. We explain each characteristic in more detail in this section.

#### 3.1 Comprehensive

EDIT is comprehensive in two ways. First, it contains more full texts of bilateral investment treaties than any other database. Second, it adopts a more inclusive definition of what constitutes an IIA to account for the diversity of investment treaties beyond BITs and integrates these treaties into a single database.

How does EDIT determine what an IIA is and whether it is included in the database? We define an IIA as a *treaty concluded under international law between two or more states, territories, or economies*,<sup>12</sup> which, in whole or in part, contains substantive obligations to protect and/or liberalize foreign investment either generally or sector specific. This notion includes bilateral investment treaties, investment chapters of preferential trade agreements and regional investment agreements (RIAs).<sup>13</sup> A few IIAs included in EDIT are not listed in UNCTAD or any other database, and consist mainly of Friendship, Commerce, and Navigation Agreements (FCNs), which have been referred to as the ‘first BITs’ (Vandeveld, 2017). At the same time, only those investment-related obligations in PTAs that fit our definition are included in EDIT; we thus exclude general framework agreements counted by UNCTAD as a treaty with investment provisions, since they do not themselves contain specific investment protection or liberalization provisions. Our definition also excludes a number of agreements that have been listed in some existing databases; for example, we do not count the ICSID Convention as an IIA.<sup>14</sup> As the purpose of EDIT is to identify past or current existence of IIA texts, terminated and renegotiated IIAs are also included in the database. We also excluded some ‘duplicate treaties’, meaning agreements that are listed as separate treaties, whereas in reality these had identical texts.<sup>15</sup>

With above definition in mind, we began compiling IIA full texts and metadata. First, we identified a list of existing public and private databases, both digital and physical, as well as government webpages and official websites of international organizations (e.g. United Nations, World Bank, OECD) where information on investment agreements can be found. In the end, we obtained a list of 146 different websites of ministries, agencies, and public institutions. UNCTAD’s IIAs Navigator proved particularly useful as the organization regularly surveys UN member states for their IIA practice and thus possesses some of the most accurate meta data and most advanced full text repository. Using these different websites, we generated a new list of agreements that would fit our definition. Where databases differed on meta-data information, we cross-checked these.

Second, based on this close to comprehensive list of agreements, we set out to collect the full texts. When an agreement’s text was unavailable on one of our 146 target websites, we contacted governments directly (through their respective ministries or agencies competent on investment issues) in order to obtain a copy of the official text of the agreement. On occasion, local personal contacts in each country were instrumental to obtain texts (Polanco Lazo *et al.*, 2018). Through this process, we were able to include treaty texts into EDIT that cannot be found in any other database.

<sup>12</sup>For the purposes of EDIT, we include territories of contested sovereignty like Kosovo, Taiwan and Palestine, among others.

<sup>13</sup>One example of RIAs is the Investment promotion and protection agreement between Japan, Republic of Korea and China (2012), available at: [www.mofa.go.jp/announce/announce/2012/5/pdfs/0513\\_01\\_01.pdf](http://www.mofa.go.jp/announce/announce/2012/5/pdfs/0513_01_01.pdf).

<sup>14</sup>EDIT excludes agreements by public and international investment insurance schemes (e.g. OPIC and MIGA), some multilateral agreements (e.g. ICSID, Mauritius Convention, TRIMs), Trade and Investment Framework Agreements (TIFAs), and Framework Agreements on Economic Cooperation.

<sup>15</sup>This is the case for treaties signed by countries that split up in various states (e.g. countries of former Yugoslavia or former Union of Soviet Socialist Republics). EDIT, however, lists all successor states which were originally part of the treaties.

**Table 1.** EDIT database

Type	N	Force	Force %	NoForce	NoForce %	Miss	Miss %	MissForce	MissForce %
<b>BITs</b>	3239	2330	71.93	909	28.06	141	4.35	9	0.27
<b>Other IIAs</b>	378	311	82.27	67	17.72	2	0.52	0	0
<b>Sum</b>	3617	2641	73.01	976	26.98	143	3.95	9	0.24

Note: Authors' calculations based on EDIT database.

EDIT strives to cover the full texts of as many IIAs as possible. Aside from the main texts, it also includes preambles, footnotes, annexes, protocols, schedules, and side-letters. When it comes to PTAs, i.e. agreements that cover investment alongside many other issues, EDIT includes investment chapters, but also all other relevant parts that relate to investment protection or liberalization such as chapters on trade in services, financial services, taxation, and general exceptions.

By May 2020, the number of IIAs included in EDIT amounted to 3617, categorized as 3239 BITs and 378 other IIAs. Of that number, we have a total of 3474 available texts meaning that we do not have access to texts of 143 IIAs that we know to exist ("Missing IIAs").<sup>16</sup> Table 1 illustrates the two types of categories of treaties found in EDIT (BITs and Other IIAs). It also reports the total count of signed treaties that are/have been in force (Force) and those that have not yet entered into force (NoForce). Amendments and replacements are also included in the available agreements. The remaining columns provide the count and percentage of treaties in force for which no text could be found (Miss) and a count of treaties that are missing and are pending (MissForce)

### 3.2. Uniformity

IIAs come in different languages. While most IIA databases focus primarily on English texts, EDIT covers a greater amount of non-English texts. Table 2 provides an overview of the different languages.

About 63% of the IIAs found in EDIT are written officially in English or have a complementary translation into that language. But some treaties are only available in another language, notably French (around 10% of the treaties) and Spanish (around 7% of the treaties), followed by Russian, German, and Arabic (each one representing around 2% of the treaties). All other languages represent less than 1% of the universe of treaties, although considered altogether represent around 10% of available texts.

To make these texts uniformly accessible and to facilitate text-as-data research across the entire corpus, we decided to provide an English version for each treaty in EDIT. All IIAs that were not found in English were translated, using machine and manual translation. The former was feasible with the help of customized machine translation software.<sup>17</sup>

We first translated all non-English treaties in French and Spanish (that constitute the largest part of non-English treaties) using the tailor-made software for IIAs. For this purpose, two machine translators were built ('MOSES'), using Joint Research Centre (JRC) texts from the EU, and 'parallel texts' of IIAs that contain Spanish/English or French/English versions in the original IIAs. Overall, 2000 sentences were drawn upon for tuning. The outcome of this machine translation provided high accuracy in short texts and common languages. MOSES was further improved with data from the UNCorpus (11.6 million sentences). MOSES translation was then

<sup>16</sup>We count 141 BITs (9 in force) and 2 other IIAs (none in force) as missing.

<sup>17</sup>The automated translation of treaties and their digital conversion was developed and applied by a research team under the supervision of: Martin Volk. Special thanks go also to Kyoko Sugisaki, Annette Rios, and Laura Mascarell.



Table 2. Most common languages

Language	N	N%
English	2268	62.70
French	361	9.98
Spanish	249	6.88
Russian	90	2.49
German	85	2.35
Arabic	71	1.96
Others	350	9.68
Missing	143	3.95

Note: N and N% indicate the number and percentage of treaties available in the original language

compared with *Google Translate* using BLEU (Bilingual Evaluation Understudy), an algorithm for evaluating the quality of text which has been machine-translated from one natural language to another with different precision (Papineni *et al.*, 2002). For all other language translations, we relied on the general machine translation engines (Google Translate, DeepL), with manual corrections by team members, research assistants, or volunteers fluent in those languages.<sup>18</sup>

### 3.3 Standardization and Machine-Readability

Aside from the translation exercise, we also did extensive manual work to convert texts into a standardized digital format. This crucial step allows EDIT users to engage in more effective search tasks, opens the door for computational content analysis techniques, and enables the annotation of IIA content.

The conversion towards machine-readability involved several steps. If original files were non-machine readable (e.g. scanned or image pdfs), we digitized the text using OCR software. This process can create errors (e.g. an ‘i’ is recognized as ‘l’). With the help of research assistants, we went through all texts to correct such errors as rigorously as possible.

Once texts were machine-readable, we standardized them into a common format and structure. We store texts in Extensible Markup Language (XML), which allows augmenting text with mark-ups for (1) meta data, (2) layout and text structures, and (3) additional content annotation.<sup>19</sup> Each XML contains information characterizing the IIA (e.g.. its parties, year of signature). This is followed by the text of the agreement, which we split into its subcomponents, i.e. its preamble, body, and annexes. Furthermore, treaty texts are split up into articles, paragraphs, and footnotes (when they exist) to capture document structure. Researchers can subsequently use this information to search exclusively in preambles or specific paragraphs. We also use this detailed document structure for the automated content annotation of IIAs.

### 3.4 Content Annotation

From researchers to policymakers, the users of EDIT are likely interested in specific IIA content features. For example, they may want to investigate all provisions dealing with ‘fair and equitable

<sup>18</sup>This task was performed by a team of researchers including the authors, as well as Dmitriy Skugarevskiy, with the invaluable help of research assistants from different countries, including Valentino Desilvestro (Italy), Azernoosh Bazrafkan (Netherlands/Iran), Lamiya Baz (Kuwait) and Mazen Hroub (Palestine), and some volunteers from all over the world (Faith Tigere from South Africa, Ngan Nguyen from Vietnam and Sebastian Espinosa from Ecuador).

<sup>19</sup>A more detailed explanation of the building of this corpus, is available at (Sugisaki *et al.*, 2016).

treatment’ or ‘most favoured nation treatment’. In contrast to existing manually coded datasets that keep full text and content information separated, EDIT’s XML format allows ‘tagging’ or labelling specific parts of the text according to content features.

We acknowledge that such a content annotation cannot be comprehensive, since it is not feasible to foresee all the specific features that a user may be interested in. Therefore, EDIT offers a first layer of content features, which registered users can expand on through their own annotation.

To implement this general layer of content features, we opted for an automated keyword-based tagging (based on Alschner and Skougarevskiy, 2016a). We systematically examined and compared treaty texts to identify recurrent expressions of core investment law concepts, e.g. the words ‘full protection and security’ and ‘complete protection and security’ indicate a protective norm of the same type. On that basis, we created a list of keywords and their corresponding content features. We then used that list to automatically tag treaty articles that contained a keyword with the content feature. To improve accuracy and distinguish between categories that share overlapping keywords, we used article titles (when they exist) to limit the scope of the search. This allowed us to distinguish a health reference in the preamble from a health-related general exception for instance. This general annotation of IIA facilitates the search for specific content features and will allow users to further refine and expand our categorization through future manual annotation.

### 3.5 A Free Academic Resource for Research

EDIT is a free of charge academic resource to support and facilitate research on international investment agreements. EDIT overcomes a limitation of some datasets that are of high-quality, but hidden behind a paywall. EDIT thereby seeks to expand access to IIA full texts. Moreover, EDIT is designed to facilitate scholarship and innovative research, in particular the computational analysis of IIAs that treats these texts as data (Gentzkow *et al.*, 2017; Grimmer and Stewart, 2013). As a result, EDIT complements, but does not seek to duplicate or compete with, other available resources on investment law.

## 4 Research Applications of EDIT

EDIT aims to facilitate investment treaty research across disciplines. In this section, we situate EDIT in current scholarship in international law, international relations, and economics, before showcasing the new descriptive insights that can be gained from it.

### 4.1 EDIT and International Law Research

We anticipate that most international lawyers will use EDIT as a searchable full text database. EDIT will provide these users with quick access to the agreements that contain the key words or content features they are interested in. The database will thereby enable practitioners and researchers to answer questions such as ‘which agreements contain references to sustainable development?’, ‘which treaty included the first counterclaims clause?’, ‘which agreements include provisions on investment facilitation?’, or ‘which treaties use identical wording on expropriation?’.

Beyond one-off searches, EDIT’s comprehensiveness and machine-readability can support a wide range of doctrinal, empirical, and computational legal research agendas. It can help investigate trends (‘How has the content of most favored nation clauses evolved over time?’);<sup>20</sup> legal policy questions (‘What treaty design features or language is correlated with losing or winning

<sup>20</sup>On the most-favoured nation (MFN) clause in BITs, see the work by Chaisse and Kirkwood (2020) and Pérez-Aznar (2017).



an investment claim?'); and large-scale content patterns ('How consistent are a country's investment agreements'). For example, researchers have used early versions of EDIT to show that treaty design is increasingly diverging rather than converging (Alschner and Skougarevskiy, 2016a); the close similarity between the Trans-Pacific Partnership (TPP) investment chapter and prior IIAs concluded by the same signatory countries (Polanco Lazo and Gómez Fiedler, 2017); the partial alignment of G20 countries treaty practice with their Guiding Principles for Global Investment Policymaking (Polanco Lazo, 2017); or that three out of four citations in investment arbitration awards reference decisions rendered under a treaty that is 50% or more dissimilar than the treaty currently interpreted (Alschner, 2020).

The perhaps greatest contribution of EDIT, however, is to help make the unknown known. Legal scholars tend to focus on *true positives* – legal acts that could happen and did happen. For example, they study the use of general exceptions in IIAs through the arbitral decisions that interpreted these exceptions (Henckels, 2018; Keene, 2017); these are illustrative, but not comprehensive. Scholars should also study *false negatives* – legal events that ought to happen, but did not. We can learn as much about the role of general exceptions in IIAs by looking at the cases where they should have been discussed – because the underlying treaty contained one such exception and the facts warranted its invocation – as by studying those decisions that did interpret them. Yet, while it is easy to spot true positives, it is hard to identify false negatives. EDIT can help in this respect. By offering a comprehensive repository of investment treaty texts, researchers can easily identify which treaties contain general exceptions and compare these to the awards rendered thereunder to investigate the whole universe of relevant observations and render the unknown known (Alschner and Hui, 2019).

#### 4.2 EDIT and International Relations and Economics Research

In international relations research, EDIT can support and weigh in on research debates related to the design determinants, evolution, and impact of investment treaties by offering a more comprehensive and machine-readable dataset.

When it comes to exploring and explaining the variation in treaty design, the BIT literature has made important advances in recent years by relying on manually coded data. There is evidence that for instance after economic downturns or civil wars, countries are more likely to sign BITs (Billing and Lugg, 2019; Simmons, 2014). Further, autocratic regimes are not shy in opting for bilateral investment treaties to bolster their political survival (Arias *et al.*, 2018; Mazumder, 2015). EDIT will allow us to tease out in more detail the types of treaties that are signed by these leaders. Scholars have also investigated what factors are responsible for specific design choices by focusing on dispute settlement clauses (Allee and Peinhardt, 2010) or on the degree of a treaty's legalization (Abbott *et al.*, 2000; Manger and Peinhardt, 2017). EDIT can complement such existing work by analysing in more depth the linguistic subtleties that vary between agreements or by comparing computational and manual text coding methods. It can also help further develop, operationalize, or refine concepts, such as 'delegation' or 'policy space' based on their varied textual expressions (Blake, 2013). By providing a more refined description of treaty variation, EDIT also opens new opportunities to study the determinants that contribute to such variation. It can be used to identify what states had a greater influence over the design and authorship of the agreement (Lugg *et al.*, 2019). The recourse to a comprehensive database allows us to detect in such events the trade-offs negotiators make within and across treaties.<sup>21</sup> In addition, by providing full texts of BIT from hitherto understudied regions, such as the Middle East and North Africa, EDIT can help tease out regional specificities. For example, using EDIT Morr Link and Yoram Haftel found that states with an Islamic Legal tradition systematically opt for more informal dispute settlement arrangements in their agreements (Link and Haftel, 2019).

<sup>21</sup>Treaty text changes are often endogenous, where one aspect is traded off with another one. See (Baccini *et al.* 2015).

EDIT also facilitates studying the evolution and diffusion of investment agreements by placing a larger dataset at the disposal of researchers. It thus becomes easier to identify where a particular treaty language first appeared. This, in turn, can support capturing and isolating certain events that trigger design changes over time for historical institutionalist scholars (Poulsen, 2019), or provide additional evidence as to the type of diffusion mechanisms (e.g. coercion, competition, learning, emulation) that could have been instrumental for the evolutionary path (Solingen, 2012). EDIT can also prove useful for estimating how treaty texts change during re-negotiations on investment treaties and what internal and external factors principally drive observed changes in treaty language (Haftel and Thompson, 2018).<sup>22</sup> In particular, EDIT creates new opportunities to study how investment arbitration impacts treaty design and how treaty design changes affect arbitral outcomes.

The data could provide a new impetus to the political-economy literature on the impact of BITs which has predominantly focused on FDI flows (Bonnitcha *et al.*, 2017; Pohl, 2018; Sauvart and Sachs, 2009; UNCTAD, 2014). Whereas the evidence on whether aggregate FDI flows have increased due to BITs remains inconclusive (Bellak, 2015; Büthe and Milner, 2009; Hallward-Driemeier, 2003), and some work finds no clear link between treaty protections and investment (Yackee, 2008, 2010), research has over time increasingly focused on the conditions under which treaties influence investment flows positively. Starting with the work by Egger and Pfaffermayr (2004), the argument has been made that the credibility of ratifying BITs and therefore the implementation stage matters. Tobin and Rose-Ackerman (2005, 2011) further find that domestic institutions and the overall investment climate matter for FDI inflows for developing countries. Allee and Peinhardt (2011) in their work provide evidence that BITs do increase FDI into countries that sign them, but only if those countries are not subsequently challenged before an arbitration tribunal (e.g. ICSID). They further observe notable losses of FDI for countries that are challenged in international arbitration. EDIT could allow us to study in more detail what type of commitments (BIT design) in conjunction with domestic factor constellations are causally related to economic effects. Following this line of research, Desbordes (2016) has found that BITs specifically granting access to ISDS and protecting foreign investors from discrimination have a large, positive, and statistically significant effect on FDI. EDIT could also prove useful for more recent work that departs from analysing aggregate FDI data and focuses more on heterogeneous effects across sectors by quantifying the effect of BITs on firm ownership (Egger *et al.*, 2019). More systematic attention to the type of sectors and the elements within global supply chains that will be most likely affected by FDI provisions can certainly complement existing case-study evidence and contribute to more knowledge about BITs economic effects.

Finally, on the impact side of investment agreements, the data will make it also easier to evaluate how commitments diffuse vertically, namely how legal texts affect the formulation of domestic Legislation (direct textual boilerplating). Also by combining investment agreement commitments in various international treaties, it can help disentangling for instance effects as part of trade agreements vs single-standing investment agreements on economic parameters ranging from investment and trade flows to wages, jobs, and technology transfers. In other words, studies can more accurately estimate the impact by types and forms of treaties and further explore where treaty commitments act as substitutes or complements.

## 5. Using EDIT to Explore the IIA Universe

In this section, we showcase possible usage of EDIT by exploring empirical questions on IIAs in law and international relations research with respect to (1) global and (2) national treaty practices as well as in relation to (3) specific treaty provisions.

<sup>22</sup>See also Thompson *et al.* (2019).

## 5.1 Global Treaty Patterns

The global structures underlying the IIA regime are of interest to lawyers, political scientists, and economists. Here, we pick two research questions with interdisciplinary significance. First, is there a global treaty model underlying the IIA universe? Second, have IIAs become less mandatory and more hortatory as treaties become longer? Such questions matter for a better understanding of the evolution, diffusion of treaty design as well as the effect of investment treaties.

### 5.1.1 Diffusion of Treaty Design: Divergent Practices or Universal Norms?

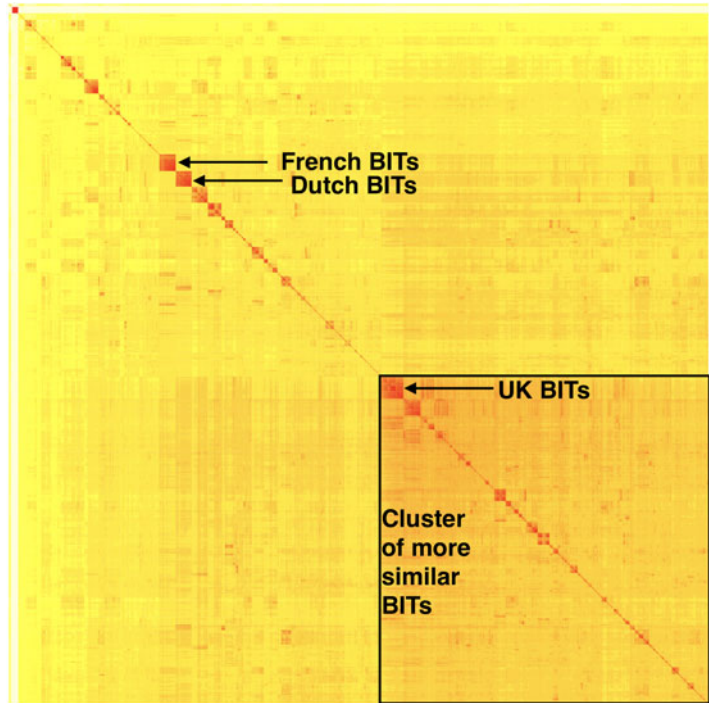
A first question is whether a specific treaty design or model dominates the IIA regime. On a spectrum between treaty design uniformity and diversity, existing research suggests that the IIA regime sits somewhere in the middle: on the one hand, scholars found that BITs tend to be highly similar within national (mostly developed) country treaty networks, but generally differ between states (Allee and Peinhardt, 2014); on the other hand, researchers have also pointed to the fact that many of these agreements, while differing in language, share a set of common norms and principles (Salacuse, 2010). By providing the most comprehensive database to date, EDIT allows to situate and trace the regime development more accurately between the extremes of national practices and universal norms and trends of convergence or divergence.

We investigate whether underlying the diverse language of national BIT programs lies a common treaty design denominator. Put differently, we want to assess to what extent a particular underlying treaty design diffused so successfully that, over time, it came to dominate the treaty network (or parts thereof). EDIT supports a myriad of different research strategies to investigate the question.

One could test existing explanations about how core features are diffusing over time and space. One could also look at imprints of model conventions and systematically test propositions, such as the one put forward by Stephan Schill, that the 1967 OECD model inspired later treaty language (Schill, 2009). However, for the purpose of illustration we let the data speak and opt here for an inductive research approach. We calculated the textual similarity of EDIT BITs and use a hierarchical clustering algorithm to group agreements by their similarity in order to find a latent model that underlies the broadest possible practice of a diverse set of states.

Figure 1 shows the results as an ordered heat map. The axes symmetrically regroup 3054 BITs and display a color-coded version of the similarity matrix: high similarity between two treaties is indicated as red, whereas yellow signifies low similarity. The red quadrangles along the diagonal axis of the heat map indicate the internally coherent national treaty networks that closely follow a national model template, on which Alschner and Skougarevskiy reported (e.g. French and Dutch BIT networks are annotated in Figure 1) (Alschner and Skougarevskiy, 2016b). In the lower right corner, however, there is a large grouping of similar worded agreements (indicated by the more reddish colouring). While the overall similarity of BITs in our dataset is 32%, the average similarity within that cluster is 40%, and 27% outside of it. The cluster of more similar BITs encompasses (parts of) the national BIT programs of several countries, including Great Britain, Israel, South Korea, India, Czechia, Hungary, Slovakia, Barbados, and others. The global/systemic analysis thus indicates that, in addition to country models, there is a latent similarity connection among a subset of BITs by a group of states.

What have these more similar BITs in common apart from their treaty language? They seem to have been inspired by the BIT design of Great Britain. The earliest BIT within that group is the 1975 Egypt–UK BIT, which is the first investment treaty the UK signed. Moreover, within the cluster, the 104 British treaties stand out for their internal consistency displaying an average of 69% of similarity. Over time, other states seem to have been inspired by the UK's treaty design. As an illustration, we compare the expropriation clause of the South Korea–Sri Lanka BIT (1980) and of the Barbados–Venezuela BIT (1994) with the same clause in the 1975 Egypt–UK BIT. They are almost perfect copies (see Table 3).



**Figure 1.** Cluster of similar agreements in global treaty network  
*Note:* The figure represents textual similarity in form of a heat map. High textual similarity is indicated as red, low textual similarity is indicated as yellow. The x and y-axis are symmetrical and regroup 3054 BITs in our dataset. The treaties are ordered through a hierarchical clustering algorithm to place similar texts together and dissimilar ones apart.

Why did the language of the UK BIT programme diffuse so successfully? Why did South Korea or Barbados not copy from German or U.S. agreements? We leave it to future researchers to find answers to these questions. We do want to highlight, however, that this analysis shows another strength of EDIT: the ability to trace diffusion even when the language of the original treaty differed. For example, in 1995 the two Baltic States Estonia and Lithuania concluded a BIT. Given their geographical proximity and close cultural and historical connection, one could have suspected that the BIT they signed was very unique in its design. Since the only publicly available version of that BIT is in Lithuanian, it would have been difficult to computationally relate its content to BITs in other languages to assess its uniqueness. EDIT's translation now enables us to situate such foreign language BITs into the wider treaty universe. It turns out that the 1995 Estonia–Lithuania BIT is not unique at all. It belongs to a family of similarly worded central eastern European agreements – to 75% it mirrors the language of the 1994 Albania–Czechia BIT and the 1994 Hungary–Mongolia BIT – which are, in turn situated in the cluster of more similar BITs inspired by the style and language of UK BITs.

### 5.1.2 Relative Normativity: Towards More Hortatory Language?

BITs have on average become longer over time. As can be seen in [Figure 2](#), the main text of BITs excluding schedules or annexes has increased from an average of around 1500 words in the 1960s to around 4000 words in the 2010s. Canadian BITs signed over the last decade even increased to an average of around 8000 words. What does this increased length mean for the design and impact of investment treaty? Have treaties become more constraining over time including more obligations or have they become less constraining by devoting treaty space to exceptions or aspirational clauses? These are important questions that developed out of international law and international relations' research programmes on rational design and legalization (Goldstein *et al.*, 2000; Koremenos *et al.*, 2001; Manger and Peinhardt, 2017).

**Table 3.** Expropriation clauses across diverse treaty parties to use language originally found in UK BITs.

Egypt–UK BIT (1975) Article 5(1)	South Korea–Sri Lanka BIT (1980) Article 7(1)	Barbados–Venezuela BIT (1994) Article 5(1)
<p>Investments of nationals or companies of either Contracting Party shall not be nationalized, expropriated or subjected to measures having effect equivalent to nationalization or expropriation (hereinafter referred to as ‘expropriation’) in the territory of the other Contracting Party except for a public purpose related to the internal needs of that Party and against prompt, adequate and effective compensation. Such compensation shall amount to the market value of the investment expropriated immediately before the expropriation itself or before there was an official Government announcement that expropriation would be effected in the future, whichever is the earlier, shall be made without delay, be effectively realizable and be freely transferable. The national or company affected shall have a right, under the law of the Contracting Party making the expropriation, to prompt review, by a judicial or other independent authority of that Party, of whether the expropriation is in conformity with domestic law and of the valuation of his or its investment in accordance with the principles set out in this paragraph.</p>	<p>Investments of nationals or companies of either Contracting Party shall not be nationalized, expropriated or subjected to measures having effect equivalent to nationalization or expropriation (hereinafter referred to as ‘expropriation’) in the territory of the other Contracting Party except for a public purpose related to the internal needs of that Party and against prompt, adequate and effective compensation. Such compensation shall amount to the market value of the investment expropriated immediately before the expropriation or impending expropriation became public knowledge, shall include interest until the date of payment, shall be made without delay, be effectively realized and be freely transferable. The national or company affected shall have a right, under the law of the Contracting Party making the expropriation, to a prompt determination of the amount of compensation either by law or by agreement between the parties and to prompt review, by a judicial or other independent authority of that Contracting Party, of his or its case and of the valuation of his or its investment in accordance with the principles set out in this paragraph.</p>	<p>Investments of nationals or companies of either Contracting Party shall not be nationalized, expropriated or subjected to measures having effect equivalent to nationalization or expropriation (hereinafter referred to as ‘expropriation’) in the territory of the other Contracting Party except for a public purpose related to the internal needs of that Party on a non-discriminatory basis and against prompt, adequate and effective compensation. Such compensation shall amount to the market value of the investment expropriated immediately before the expropriation or before the impending expropriation became public knowledge, whichever is the earlier, shall include interest at a normal commercial rate until the date of payment, shall be made without delay, be effectively realizable and be freely transferable. The national or company affected shall have a right, under the law of the Contracting Party making the expropriation, to prompt review, by a judicial or other independent authority of that Party, of his or its case and of the valuation of his or its investment in accordance with the principles set out in this paragraph.</p>



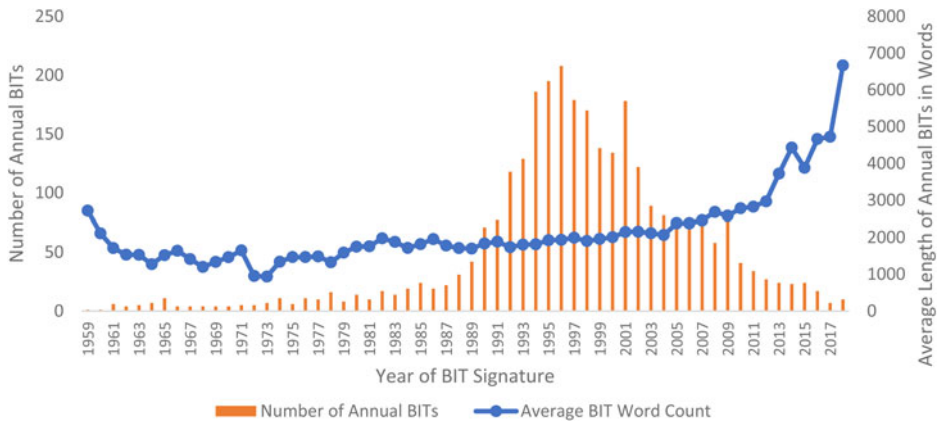


Figure 2. Expansion of BITs and increase in treaty length

One way of answering this question of relative normativity (or degree of obligation or ‘bindingness’ of law) is by investigating whether longer BITs contain more mandatory commitments (‘shall’) or hortatory expectations (‘should’, ‘may’). If the share of mandatory clauses has remained constant or even increased, this would suggest that longer BITs impose more commitments on its signatories. If, however, the share of mandatory clauses has declined and, conversely, the share of more hortatory language has increased, this would suggest that BITs are increasingly used to set forth best-effort clauses and the average level of obligation decreases. EDIT enables researchers to track such relative normativity over time.<sup>23</sup>

We have investigated mandatory language – must, shall, will – as well as hortatory language – may, should, endeavour, strive – in order to discern trends over time. In absolute terms, both mandatory and hortatory language in BITs has expanded over time. BITs signed in the 1960s contained on average 90 mandatory words and seven hortatory words per treaty; BITs signed in the 2010s contain 140 mandatory words and 19 hortatory words. However, this increase is in part due to the fact that treaties got longer.

When we look at mandatory and hortatory language in relative terms (occurrence per 100 words), then the impression changes, as can be seen in Figure 3. Mandatory words have decreased over time. That is, less treaty space is today devoted to setting hard, normative commitments. At the same time, the relative inclusion of hortatory language has barely increased. As a result, BITs today are less about setting forth strong commitments; yet, neither has the focus shifted to aspirational clauses.

At the same time, in absolute terms, hortatory clauses have become more common. One frequent critique in this respect is that while traditional investment protection obligations are typically framed in mandatory language, more recent non-investment protection obligations, such as commitments to protect the environment, labour rights, or to promote corporate social responsibility, are consistently framed in hortatory term (Chi, 2015; Morin *et al.*, 2018). To test this empirically, we investigated all clauses on environmental protection, labour rights, and corporate social responsibility. We found that, indeed, all of them are couched, at least in part, in hortatory language. As calls for binding investor obligations to respect human rights intensify, future iterations of this analysis will show whether these hortatory clauses ‘harden’ into mandatory investor obligations in new IIAs (Choudhry, 2020). Textual analysis based on signalling words can quickly identify the degree of normativity of different provisions and describe patterns over time.

<sup>23</sup>This research question is inspired by and operationalizes the seminal work of Weil (1983)



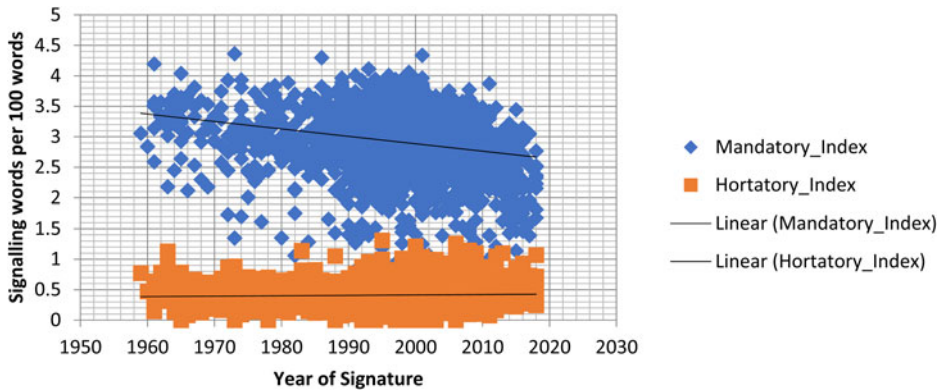


Figure 3. Mandatory and hortatory language of IIAs over time

## 5.2 National Treaty Patterns

We now turn from global to state-level patterns. Here, EDIT allows researchers, for the first time, (1) to systematically explore non-English treaty networks and (2) to investigate in more depth why some countries are rule-takers and others rule-makers.

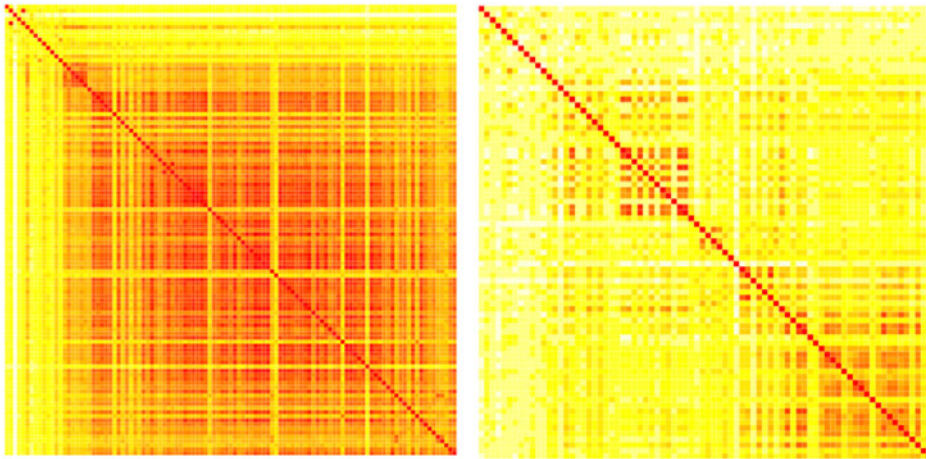
### 5.2.1 Consistency and Innovation: Exploring Non-English Treaty Networks

Alschner and Skougarevskiy used a dataset of predominantly English agreements to track consistency and innovation in national BIT networks (Alschner and Skougarevskiy, 2016b). EDIT now allows us to explore countries with predominantly non-English treaty networks. Only seven English-language BITs of France, for example, are included in the original mapping by Alschner and Skougarevskiy; EDIT now contains the text of 113 French BITs. Similarly, only 23 Russian BITs up to 2006 were previously available; EDIT now includes 82 BITs up to 2016. Figure 4 visualizes the textual similarity of the respective treaty networks. France's BITs concluded between 1963 and 2007 display an average similarity of 55%, placing it among the most consistent treaty networks in the world. Russia, in contrast, has a comparatively inconsistent treaty network. Its BITs concluded between 1989 and 2016 on average have only 45% of their text in common. However, Russia's heat map illustrates an evolution in the country's BIT policy. In 1998, Russia seems to have adopted a new BIT model, which it followed particularly consistently in the 2000s and 2010s resulting in a new generation of more similar agreements with an average similarity of 64% color-coded in red in the lower right corner.

By allowing us to track the changes to national investment treaty programs, EDIT can monitor responses to current criticism against the IIA regime. While some countries have denounced BITs or stopped concluding new ones, other states have drastically departed from their earlier investment agreements to conclude new BITs that remedy shortcomings in earlier BITs.

One of these countries is Colombia. Up to 2006, Colombia concluded relatively short and simple agreements that focused exclusively on the protection of investors. Although the country had not faced any investment claims, it decided to revise its IIA practice in 2006 in response to global developments in ISDS. As a Colombian representative explained, its revised 2006 model reflected 'careful consideration of [investor-state arbitration] awards from that time, interpretations from the NAFTA Free Trade Commission and consultations with UNCTAD officials'.<sup>24</sup> The IIAs Colombia began to sign from 2007 onwards then embodied a more detailed and elaborate balance

<sup>24</sup>Statement of Colombia, United Nations Conference on Trade and Development (UNCTAD) World Investment Forum, Geneva, 14 October 2014. Available at: <https://worldinvestmentforum.unctad.org/wp-content/uploads/2014/10/Vargas-Saldarriaga.pdf>.



**Figure 4.** Consistency (textual similarity) heat maps of the BIT networks of France (left) and Russia (right). Axes list BITs chronologically by date of signature (earliest treaties upper left, latest treaties lower right).

of investment protection commitments, on the one hand, and public policy flexibilities, on the other hand.<sup>25</sup> Figure 5 neatly traces this shift. While its newest generation of BITs does not display heightened consistency, the white and yellow tiles in the upper right and lower left corner indicate that the new generation of agreements is extremely dissimilar from earlier Colombian treaties. As more states revise their approach to BITs, EDIT can track how radical the departure from the past has been.

### 5.2.2 Negotiation Success: Rule-Takers versus Rule-Makers?

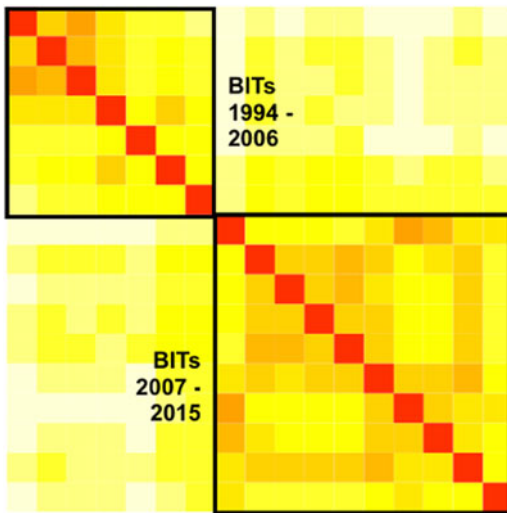
Previous research suggests that the consistency of BIT language over time is a proxy for the negotiation success of states (Allee and Peinhardt, 2014; Alschner and Skougarevskiy, 2016b; Berge and Stiansen, 2016). Powerful capital exporting countries tend to reproduce highly coherent treaty language across their negotiated agreements. They thus tend to be the system's rule-makers. Vice versa, less developed capital importing countries tend to have a patchwork of relatively inconsistent agreements. They appear to sign on to the language proposed by their more powerful negotiation partners and are thus the system's rule-takers. In short, a country's consistency of BIT language correlates strongly with economic power expressed as GDP per capita (Alschner and Skougarevskiy, 2016b).

However, some similarly situated countries drastically differ in their treaty consistency. Hence, factors other than economic power need to be considered in the literature to understand negotiation success. Besides economic power, past experience in negotiations, geographic location, or shared historical trajectories might predict design similarity.

Let us consider the Czech Republic and Poland for example. Poland has four times the population of the Czech Republic and therefore a larger GDP (although its GDP per capita is lower). This would suggest that Poland has a bit more negotiation leverage. Moreover, Poland was one of the first countries east of the Iron Curtain to conclude investment treaties starting in the late 1980s and acquired considerable negotiation expertise even prior to the end of the Warsaw Pact. The Czech Republic only entered the BIT universe in the 1990s and started to seriously pursue BITs only after the break-up of Czechoslovakia in 1992.<sup>26</sup> Hence, based on economic power and experience, one could expect that Poland and not the Czech Republic developed a more

<sup>25</sup>Ibid.

<sup>26</sup>Other dimensions (e.g. geography, new democracies) might predict similar approaches.



**Figure 5.** Consistency (textual similarity) heat map of the BIT network of Colombia. Axes list BITs chronologically by date of signature (earliest treaties upper left, latest treaties lower right). Framed quadrangles help highlight the shift in Colombia's BIT practice between distinct periods when treaties were more internally consistent.

consistent BIT regime. Yet the opposite is true. While the Czech Republic achieved a highly consistent treaty network (51% average similarity), Poland was less successful (43% average similarity). This difference in consistency persists (and increases) when we control for selected treaty partner characteristics. Below, we illustrate examples where the counterparts were economically less powerful. In other words, even in instances of potential rule-making power, Poland followed a much less consistent approach than the Czech Republic. [Table 4](#) lists the relative similarity of the BITs the two states signed with Latvia, Romania, Vietnam, and Mongolia in the 1990s. While in the case of the Czech Republic these four BITs resulted in overall highly similar agreements, the BITs Poland signed differed significantly overall ([Table 3](#)).

What can be observed is a puzzle. Why do some states negotiate highly consistent agreements while others fail to do so in similar situations? Whereas relative economic power is an important indicator, in this case it cannot explain the variation we observe. Future research can help us better understand when states opt for more or less consistency, through which agreements they might change their trajectory and how and when they might transform from rule-takers to rule-makers.

### 5.3 Specific Clauses: Changing Preambles

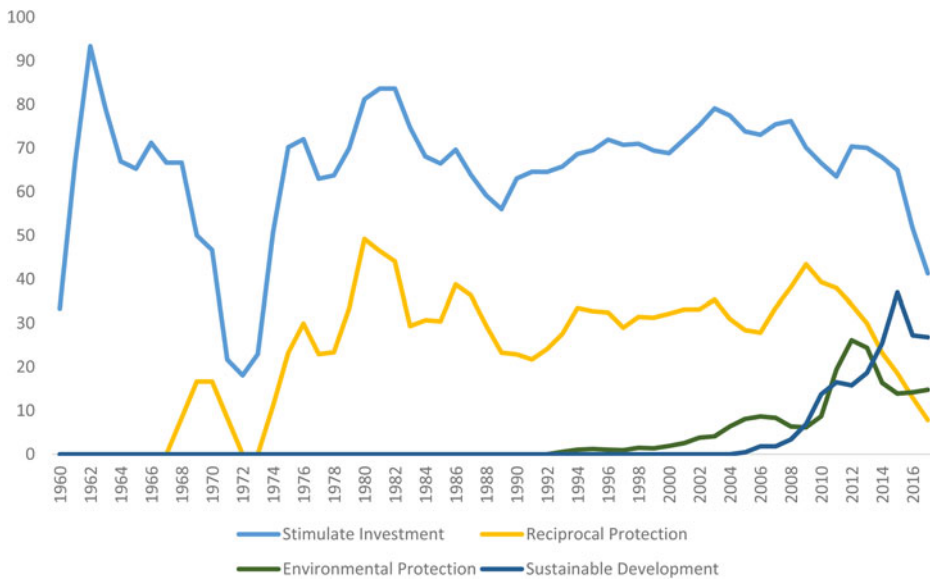
Computational analysis can also be conducted on the level of specific clauses, given that EDIT provides for structured treaty texts that can be broken down into clauses. Here we exemplify this procedure by assessing the preambles in all BITs.

Preambles do not create rights or obligations, but they set out the goals, aspirations, and reasons for entering into BITs, which is important to understand the regime's shared principles (Ruggie, 1982). A textual analysis of preambles offers a window into how these contracting objectives have evolved. Preambles have become more extensive over time from an average length of 35 words in the 1960s to an average of 55 words in 2015. Moreover, new considerations have been added to preambles. [Figure 6](#) tracks the share of annually concluded BITs over time that contain a set of key words that illustrate the gradual inclusion of new considerations into BITs.

First, although BITs were substantively concerned with the protection of foreign investment rather than its promotion or liberalization, BITs included the stimulation of capital flows as the principal purpose of the treaties from the outset. Already the 1959 Germany–Pakistan BIT expressed the hope that the treaty would ‘promote investment’. The majority of subsequent BITs similarly indicated their expectation that the BIT would ‘stimulate’ investment flows. In

**Table 4.** Textual overlap of BITs signed by the Czech Republic and Poland in the 1990s with selected countries (percentage indicates textual overlap)

BITs of Czech Republic	Latvia (1993 BIT)	Romania (1994 BIT)	Vietnam (1997 BIT)	Mongolia (1998 BIT)
Latvia (1993 BIT)	–	90%	83%	84%
Romania (1994 BIT)	90%	–	85%	86%
Vietnam (1997 BIT)	83%	85%	–	90%
Mongolia (1998 BIT)	84%	86%	90%	–
BITs of Poland	Latvia (1993 BIT)	Romania (1994 BIT)	Vietnam (1994 BIT)	Mongolia (1995 BIT)
Latvia (1993 BIT)	–	54%	46%	54%
Romania (1994 BIT)	54%	–	–	46%
Vietnam (1994 BIT)	46%	54%	–	59%
Mongolia (1994 BIT)	54%	46%	59%	–



**Figure 6.** Share of annually concluded BITs with key terms in Preamble (3-year moving average)

recent BITs, however, the reference is found less frequently. This speaks to the work of Lauge Poulsen who claims governments in developing countries typically overestimated the economic benefits of investment treaties and practically ignored their risks (Poulsen, 2015).

Second, in the late 1960s, BIT preambles began to frequently refer to the objective of ‘reciprocal protection.’ Such preambular wording seems to have originated on the behest of developing countries, starting consistently with Indonesia through the 1969 Indonesia–Norway and the 1970 Indonesia–Belgium BITs.<sup>27</sup> Although BITs were formally reciprocal from the outset, earlier

<sup>27</sup>A reference to ‘reciprocal contractual protection’ was already inserted in the 1962 Turkey–Germany BIT.

preambles, following the 1962 OECD Draft Convention on Foreign Property, gave the impression of a one way-street: ‘promoting the flow of capital [from the North] for economic activity and development [in the South].’<sup>28</sup> While, de facto, investment flows in the 1960s and 1970s were likely more uni- rather than bidirectional, the emphasis on reciprocity by developing states underscored sovereign equality at a time of decolonization.

Third, references to ‘environmental protection’ began to appear in BITs in 1994. This development was spearheaded by the United States, which, starting with the 1994 USA–Georgia BIT, began to assert that investment protection and promotion could be achieved without relaxing health, safety, or environmental protection. Around 2000, this wording was picked up in a series of BITs by Finland (starting with its 2000 BIT with Bosnia and Herzegovina) and by Trinidad and Tobago (starting with its 1999 BIT with Cuba) before becoming more widely adopted in the 2000s and 2010s.

Fourth, explicit mentions of ‘sustainable development’ as one of the BIT goals appeared first in the 2006 BIT between Canada and Peru. A year earlier the International Institute for Sustainable Development (IISD) had published its model agreements on investment for sustainable development. In 2016, two thirds of BITs concluded that year contain references to sustainable development.

With the above snap-shot on the preambles, EDIT allows us to detect shared principles and norms that are introduced in the treaty network and also locate first movers (or rule-makers) in the context of competing norms that have shaped the investment regime.

## 6. Conclusions

The recent politicization of investment treaties has increased public and scholarly interest into the investment regime while the demand for more systematic data has gone up as well. This article has introduced EDIT and has provided some illustrations for its potential use and empirical application in fields of international law and international relations. Further, we think that the data can be of significant interest to researchers applying network analysis and econometric modelling for studying evolution and impact of the investment regime. Using advanced methods across disciplines, EDIT can contribute to answering questions with promising interdisciplinary applications. In conclusion, EDIT aims to stimulate the application of emerging new methods of text retrieval and analyses and provide an empirical backbone for engaging with both older and emerging debates in international relations, international law, and international economics about the role of investment treaties.

**Acknowledgments.** We acknowledge funding from the Swiss Network for International Studies (SNIS), the World Bank and the John R. Evans Leaders Grant, Canada Foundation for Innovation. We wish to thank Dmitriy Skougarevskiy, the SNIS project team and research assistants based at University of Ottawa and the World Trade Institute for their contribution to the database.

## References

- Abbott KW, Keohane RO, Moravcsik A, Slaughter A-M, and Snidal D (2000) The Concept of Legalization. *International Organization* 54(3), 401–419.
- Allee T and Peinhardt C (2010) Delegating Differences: Bilateral Investment Treaties and Bargaining Over Dispute Resolution Provisions. *International Studies Quarterly* 54(1), 1–26.
- Allee T and Peinhardt C (2011) Contingent Credibility: The Impact of Investment Treaty Violations on Foreign Direct Investment. *International Organization* 65(3), 401–432.
- Allee T and Peinhardt C (2014) Evaluating Three Explanations for the Design of Bilateral Investment Treaties. *World Politics* 66(1), 47–87.

<sup>28</sup>Preamble of the 1962 OECD Draft Convention on the Protection of Foreign Property, available at: <https://www.oecd.org/daf/inv/internationalinvestmentagreements/39286571.pdf>.



- Alschner W** (2020) Ensuring Correctness or Promoting Consistency? Tracking Policy Priorities in Investment Arbitration through Large-Scale Citation Analysis. In Fauchald OK, Behn D, and Langford M (eds.), *Empirical Perspectives on Investment Treaty Arbitration*, Cambridge University Press.
- Alschner W and Hui K** (2019) Missing in Action: General Public Policy Exceptions in Investment Treaties. In Sachs LE, Coleman J, and Johnson L (eds.), *Yearbook on International Investment Law and Policy 2018*, Rochester: Oxford University Press.
- Alschner W, Pauwelyn J, and Puig S** (2017) The Data-Driven Future of International Economic Law. *Journal of International Economic Law* 20(2), 217–231.
- Alschner W and Skougarevskiy D** (2016a) Convergence and Divergence in the Investment Treaty Universe – Scoping the Potential for Multilateral Consolidation. *Trade, Law and Development* 8(2), 331–362.
- Alschner W and Skougarevskiy D** (2016b) Mapping the Universe of International Investment Agreements. *Journal of International Economic Law* 19(8), 561–588.
- Arias E, Hollyer JR, and Rosendorff BP** (2018) Cooperative Autocracies: Leader Survival, Creditworthiness, and Bilateral Investment Treaties. *American Journal of Political Science* 62(4), 905–921.
- Baccini L, Dür A, and Elsig M** (2015) The Politics of Trade Agreement Design: Revisiting the Depth–Flexibility Nexus. *International Studies Quarterly* 59(4), 765–775.
- Bellak C** (2015) Economic Impact of Investment Agreements. Department of Economics Working Paper Series. WU Vienna University of Economics and Business, Vienna, 200(200), [www.wu.ac.at/economics/forschung/wp/](http://www.wu.ac.at/economics/forschung/wp/)
- Berge TL and Stiansen Ø** (2016) Negotiating BITs with Models. The Power of Expertise. PluriCourts Research Paper No. 16-13.
- Billing T and Lugg AD** (2019) Conflicted Capital: The Effect of Civil Conflict on Patterns of BIT Signing. *Journal of Conflict Resolution* 63(2), 373–404.
- Blake DJ** (2013) Thinking Ahead: Government Time Horizons and the Legalization of International Investment Agreements. *International Organization* 67(4), 797–827.
- Bonnitcha J, Poulsen LNS, and Waibel M** (2017) *The Political Economy of the Investment Treaty Regime*, Oxford: Oxford University Press.
- Brown C (ed.)** (2011) *Evolution in Investment Treaty Law and Arbitration*, Cambridge: Cambridge University Press.
- Büthe T and Milner HV** (2009) Bilateral Investment Treaties and Foreign Direct Investment: A Political Analysis. In K. P. Sauvant and L. E. Sachs (eds.), *The Effect of Treaties on Foreign Direct Investment: Bilateral Investment Treaties, Double Taxation Treaties, and Investment Flows*, Oxford: Oxford University Press, pp. 171–224.
- Chaisse J and Bellak C** (2015) Navigating the Expanding Universe of International Treaties on Foreign Investment: Creation and Use of a Critical Index. *Journal of International Economic Law* 18(1), 79–115.
- Chaisse J and Kirkwood J** (2020) Chinese Puzzle: Anatomy of the (invisible) Belt and Road Investment Treaty. *Journal of International Economic Law* 23(1), 245–269.
- Chi M** (2015) The ‘Greenization’ of Chinese Bits: An Empirical Study of the Environmental Provisions in Chinese Bits and its Implications for China’s Future Bit-Making. *Journal of International Economic Law* 18(3), 511–542.
- Choudhury B** (2020) Investor Obligations for Human Rights. *ICSID Review* (Forthcoming)
- Desbordes R** (2016) A Granular Approach to the Effects of Bilateral Investment Treaties and Regional Trade Investment Agreements on Foreign Direct Investment, Presented at the Asian Economic Integration Report 2016 theme chapter on ‘What Drives Foreign Direct Investment in Asia and the Pacific?’ Background Paper, [https://aric.adb.org/pdf/events/aced2016/paper\\_rodolphedesbordes.pdf](https://aric.adb.org/pdf/events/aced2016/paper_rodolphedesbordes.pdf)
- Dür A, Baccini L, and Elsig M** (2014) The Design of International Trade Agreements: Introducing a New Dataset. *Review of International Organizations* 9(3), 353–375.
- Egger P, Erhardt K, and Masllorens G** (2019) The Impact of Investment Treaties on Firm Ownership, Presented at the European Trade Study Group, Bern, [www.etsg.org/ETSG2019/papers/110.pdf](http://www.etsg.org/ETSG2019/papers/110.pdf)
- Egger P and Pfaffermayr M** (2004) The Impact of Bilateral Investment Treaties on Foreign Direct Investment. *Journal of Comparative Economics* 32(4), 788–804.
- Gentzkow, M, Kelly, B T and Taddy, M.** (2017) *Text as Data*, National Bureau of Economic Research, [www.nber.org/papers/w23276.pdf](http://www.nber.org/papers/w23276.pdf)
- Goldstein J, Kahler M, Keohane RO, and Slaughter A-M** (2000) Introduction: Legalization and World Politics. *International Organization* 54(3), 385–399.
- Grimmer J and Stewart BM** (2013) Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis* 21(3), 267–297. doi:10.1093/pan/mps028.
- Haftel YZ and Thompson A** (2018) When Do States Renegotiate Investment Agreements? The Impact of Arbitration. *The Review of International Organizations* 13(1), 25–48.
- Hallward-Driemeier M** (2003) *Do Bilateral Investment Treaties Attract Foreign Direct Investment? Only a Bit? and They Could Bite*, Vol. 3121, The World Bank, <http://documents.worldbank.org/curated/en/113541468761706209/pdf/multi0page.pdf>



- Henckels C** (2018) Scope Limitation or Affirmative Defence? The Purpose and Role of Investment Treaty Exception Clauses. L. Bartels and F. Paddeu (eds.), *Exceptions and Defences in International Law*. Oxford University Press.
- Keene A** (2017) The Incorporation and Interpretation of WTO-Style Environmental Exceptions in International Investment Agreements. *The Journal of World Investment and Trade* 18(1), 62–99.
- Koremenos B, Lipson C, and Snidal D** (2001) The Rational Design of International Institutions. *International Organization* 55(4), 761–799.
- Link M and Haftel YZ** (2019) Islamic Legal Tradition and the Choice of Investment Arbitration Forums. *Review of International Political Economy*. doi: 10.1080/09692290.2019.1679223.
- Lugg A, Lund N, Allee T, and Elsig M** (2019) Determining the Authorship of Preferential Trade Agreements: A New Technique Using a Supervised Author Topic Model v2.0., Presented at the The Political Economy of International Organizations (PEIO), University of Salzburg, Austria. [www.peio.me/wp-content/uploads/2019/01/PEIO12\\_paper\\_108.pdf](http://www.peio.me/wp-content/uploads/2019/01/PEIO12_paper_108.pdf).
- Manger MS and Peinhardt C** (2017) Learning and the Precision of International Investment Agreements. *International Interactions* 43(6), 920–940.
- Mazumder S** (2015) Can I Stay a BIT Longer? The Effect of Bilateral Investment Treaties on Political Survival. *The Review of International Organizations* 4(11), 477–521.
- Morin J-F, Dür A, and Lechner L** (2018) Mapping the Trade and Environment Nexus: Insights from a New Data Set. *Global Environmental Politics* 18(1), 122–139.
- Papineni K, Roukos S, Ward T, and Zhu W-J** (2002) BLEU: A Method for Automatic Evaluation of Machine Translation. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. Philadelphia, PA: Association for Computational Linguistics, pp. 311–318.
- Pérez-Aznar F** (2017) The Use of Most-Favoured-Nation Clauses to Import Substantive Treaty Provisions in International Investment Agreements. *Journal of International Economic Law* 20(4), 777–805.
- Pohl J** (2018) Societal Benefits and Costs of International Investment Agreements. OECD Working Papers on International Investment, 2018/01. doi:10.1787/e5f85c3d-en.
- Polanco Lazo R** (2017) G20 International Investment Agreements and Guiding Principles for Global Investment Policymaking. <http://e15initiative.org/publications/g20-international-investment-agreements-and-guiding-principles-for-global-investment-policymaking/>.
- Polanco Lazo R, Desilvestro V, and Bazrafkan A** (2018) Missing Investment Treaties. *Journal of International Economic Law* 21(3), 703–731.
- Polanco Lazo R and Gómez Fiedler S** (2017) A Requiem for the Trans-Pacific Partnership: Something New, Something Old and Something Borrowed? *Melbourne Journal of International Law* 18(2), 1–50.
- Poulsen LNS** (2015) *Bounded Rationality and Economic Diplomacy: The Politics of Investment Treaties in Developing Countries*. Cambridge University Press.
- Poulsen LNS** (2019) Beyond Credible Commitments: (Investment) Treaties as Focal Points. *International Studies Quarterly*. doi:10.1093/isq/sqz079.
- Ruggie JG** (1982) International Regimes, Transactions, and Change: Embedded Liberalism in the Postwar Economic Order. *International Organization* 36(2), 379–415.
- Salacuse JW** (2010) The Emerging Global Regime for Investment. *Harvard International Law Journal* 51, 427–553.
- Sauvant KP and Sachs LE** (eds.) (2009) *The Effect of Treaties on Foreign Direct Investment: Bilateral Investment Treaties, Double Taxation Treaties, and Investment Flows*. Oxford University Press.
- Schill SW** (2009) *The Multilateralization of International Investment Law*. Cambridge University Press.
- Simmons BA** (2014) Bargaining over BITs, Arbitrating Awards: The Regime for Protection and Promotion of International Investment. *World Politics* 66(1), 12–46.
- Solingen E** (2012) Of Dominoes and Firewalls: The Domestic, Regional, and Global Politics of International Diffusion. *International Studies Quarterly* 56(4), 631–644.
- Sugisaki K, Volk M, Polanco R, Alschner W, and Skougarevskiy D** (2016) Building a Corpus of Multi-lingual and Multi-format International Investment Agreements. In 29th International Conference on Legal Knowledge and Information Systems (Jurix), Nice, December 2016, Nice. doi:10.3233/978-1-61499-726-9-203.
- Thompson A, Broude T, and Haftel YZ** (2019) Once Bitten, Twice Shy? Investment Disputes, State Sovereignty, and Change in Treaty Design. *International Organization* 73(4), 859–880.
- Tobin J and Rose-Ackerman S** (2005) Foreign Direct Investment and the Business Environment in Developing Countries: The Impact of Bilateral Investment Treaties. *Yale Law and Economics Research Paper* 293, 1–51.
- Tobin J and Rose-Ackerman S** (2011) When BITs Have some Bite: The Political-Economic Environment for Bilateral Investment Treaties. *The Review of International Organizations* 6(1), 1–32.
- UNCTAD** (2014) The Impact of International Investment Agreements on Foreign Direct Investment: An Overview of Empirical Studies 1998–2014. IIA Issues Note, <http://investmentpolicyhub.unctad.org/Upload/Documents/unctadweb-diae-pcb-2014-Sep%2024.pdf>.
- UNCTAD** (2015) *World Investment Report 2015. Reforming International Investment Governance*. New York and Geneva: United Nations.

UNCTAD (2019) *World Investment Report 2019. Special Economic Zones*. New York and Geneva: United Nations.

Vanderveelde KJ (2017) *The First Bilateral Investment Treaties: US Postwar Friendship, Commerce and Navigation Treaties*. Oxford University Press.

Weil P (1983) Towards Relative Normativity in International Law? *American Journal of International Law* 77(3), 413–442.

Yackee JW (2008) Bilateral Investment Treaties, Credible Commitment, and the Rule of (International) Law: Do BITs Promote Foreign Direct Investment? *Law and Society Review* 42(4), 805–832.

Yackee JW (2010) Do Bilateral Investment Treaties Promote Foreign Direct Investment? Some Hints from Alternative Evidence. *Virginia Journal of International Law* 51, 397–442.