

5. Working Paper

This section of the report explains in general terms, what has been accomplished by the end of the project “Diffusion of International Law: A Textual Analysis of International Investment Agreements (IIAs)” (hereinafter “the Project”), following the SNIS rules and procedure for submitting reports and working papers (as amended 7 April 2017).

In the following sections, you will find a description of the current state of the research, including the methods employed, problems encountered and what actions were taken or will follow, results and practical application of the research, publications and collaboration between partners.

5.1. State of the Art

Currently there is no comprehensive database of the texts that includes investment provisions found in different international investment agreements (IIAs). The most complete database is administered by UNCTAD [<http://investmentpolicyhub.unctad.org/IIA>].

The language of the treaties varies as they are found in the common or official language of the contracting parties with roughly half of the agreements in English. Many IIAs are found exclusively in French and Spanish, and some only in Russian, Arabic, Portuguese and German.

Other existing databases provide less information, like the United Nations Treaty Collection [<https://treaties.un.org/>], and ICSID's “Investment Laws of the World” (only in printed loose-leaf). Kluwer Arbitration “BITS” [<http://www.kluwerarbitration.com/CommonUI/BITS-countries.aspx>] features around 1400 treaties supplied by the Penn State Institute of Arbitration. Similarly, Oxford's “Investment Claims” [<http://oxia.oup.com/#>] includes around 1200 BITS. However, both contain agreements only in English-language. Other databases are more focused on decisions of ISDS tribunals; Investment Treaty Arbitration – ITA [<http://www.italaw.com/investment-treaties>], and the Investor-Law Guide – ISLG [<http://www.investorstatelawguide.com/>]. Some of the databases require a paid subscription (e.g., Oxford Investment Claims and ISLG). Few of these databases contain the relevant parts of PTAs as IIAs (UNCTAD, ITA, ISLG), but even in those cases their inclusion is partial and incomplete.

The literature on IIAs is mainly devoted to ISDS case law, and has been limited to certain specific features of investment treaties. These include dispute settlement provisions [Pohl et al. 2012], issues related to shareholders claims [Gaukrodger 2014], sustainable development concerns [Gordon et al. 2014] and temporal validity of investment claims [Pohl 2013]. Another branch of the literature has been devoted to examine the effects that IIAs may have (or not) on the inflows of foreign direct investment (FDI). However, the evidence is not conclusive. Some studies show that IIAs have a positive impact as they minimize political risk [Kerner & Lawrence, 2012; Neumayer & Spess, 2005], others suggest that these treaties have a modest or limited impact [Salacuse & Sullivan, 2005; Bergera et al., 2010], and some

provide evidence that IIAs have no impact or even a negative one [Hallward-Driemaier M., 2003; Poulsen, 2010].

Some databases have coded specific provisions in investment treaties and trade agreements. In BITSEL, eleven important elements of drafting variation have been coded of a universe of more than 1,500 BITs and more than 100 PTAs with investment chapter [Chaisse & Bellak, 2015]. In DESTA, the coding includes seven of the most relevant provisions with respect to non-discrimination, market access and ISDS, from approximately 790 PTAs signed period between 1947 and 2010 [Baccini, Dür & Elsig 2015]. PLURICOURTS is coding investment treaties invoked in investor-State arbitration cases as a part of a larger project on an investment treaty arbitration database [Pluricourts, 2015]. UNCTAD has recently launched an “IIA University Mapping Project” and categorizes more than 120 variations on key provisions of 2573 treaties found in its database [UNCTAD, 2015].

However, these databases do not compare the actual texts of the treaties. In thousands of IIAs, even “standard” provisions in investment treaties – like national treatment or protection against expropriation – admit several variations that are not captured by existing coding approaches. Usually coding is not performed in provisions that are very specific of IIAs, like the ones granting pre-establishment rights (when included in IIAs) that are usually subject to country-specific exceptions (e.g. PTAs annexes).

5.2. Demonstration of the Research Gap

Investment treaties come in different shapes. As mentioned, the large majority of them are BITs, but in the last decade investment provisions have been increasingly included in PTAs. A first innovation of this new database is the integration and merging of all relevant texts in one database. While some databases include both types of agreements, PTA’s texts are not harmonized with BITs (some are duplicated) and the text of several PTAs is not complete in relevant parts referred to investment (particularly with respect to taxation and financial measures).

Investment treaty databases are also incomplete. As mentioned, the most comprehensive database is the one administered by UNCTAD. This database includes information of a total of 3608 IIAs, but with only 3134 available texts, with 474 missing texts, including 455 BITs (165 in force) and 19 TIPs (7 in force). The total number of IIAs we have included in EDIT is 3523, with 3165 BITs and 358 other IIAs (or “Treaties with Investment Provisions” – TIPs). From that total number of treaties, at the moment of the official end of the project it was not possible to find the text of 185 IIAs (“Missing IIAs”), including 179 BITs (only 22 in force) and 6 Other IIAs (only 3 in force). The total number of available texts in EDIT is therefore 3338 treaties. A second innovation of our database is therefore the creation of the most systematic database.

Table 1 illustrates the composition of EDIT. The column N reports the total count of signed treaties, that are/have been in force and NoForce stands for treaties that have not (yet) entered into force. The remaining columns provide the count and percentage of treaties for which no text could be found. Overall the texts of 185 BITs are still missing, with the vast majority of them being treaties not in force. Only 25 treaties that are currently in force are still missing. Few IIAs included in EDIT are not listed in UNCTAD or any other database. Similarly, a small number of available treaties is incomplete, being largely PTAs or RIAs, which only are

included in databases with their investment chapter, omitting relevant parts of the same agreement that are applicable to investment like chapters on trade in services, financial services and general exceptions – including taxation issues.

Table 1: EDIT database

Type	N	Force	Force %	NoForce	NoForce %	Miss	Miss %	MissForce	MissForce %
BIT	3,163	2,527	79.89	636	20.11	179	5.79	22	0.70
FCN	46	46	100	0	0	0	0	0	0
OIA	312	262	83.97	50	16.03	6	2.56	5	1.60
Sum	3,521	2,835	80.52	686	19.48	185	5.42	27	0.77

This table shows the composition of the EDIT database.

The search of IIAs was performed looking for the official texts of agreements concluded before December 2016, regardless of them being in force or not, although the ratification status was recorded for the purposes of BITs determinants.

As the purpose of the study is to identify the existence of a text for negotiated IIAs, terminated and renegotiated IIAs were also included in the sample. When the text was not available in English, the analysis of IIAs was done by researchers knowledgeable on both investment law and the language of the agreement.

Following the definition of IIAs used in the project, some agreements that have been excluded from existing IIAs databases were added to EDIT: a total of 46 Friendship, Commerce and Navigation Agreements (FCNs). Few other IIAs that are not currently listed in UNCTAD (a total of 8 agreements) were also added, as they corresponded to 4 BITs and 4 PTAs with investment provisions.

It is also important to note that IIAs have been subdivided at article level (and annotated at a paragraph level). The total number of articles is 67'997 in 3'366 treaties.

Another challenge with the current collection of investment treaties is the language. Today, databases include treaties mainly in their original language of negotiation. Recent research has gathered a total number of 1623 treaties available in English [Alschner & Skougarevskiy, 2015]. Given that the universe of IIAs exceeds 3,300 agreements, a large number of treaties are not included due to language constraints, making comparisons of texts with similar provisions in IIAs in another language tricky. For example, entering “national treatment” as search criteria in in the UNCTAD database generates 1636 results. Searching for the equivalent in French (“*traitement national*”) and Spanish (“*trato nacional*”) gives only 275 and 181 results, respectively, and only a fraction of them are at the same time included in the English’ research results. A third innovation of our project is therefore that it compiles treaties in a single language (English) in one database. This is feasible with the help of machine translation.

Another challenge with investment treaties data analyses is the format. While some of the abovementioned databases provide full texts for a number of treaties, several IIAs are in image or photographic presentation and therefore it is necessary to make an electronic conversion into machine-encoded text by using Optical Character Recognition (OCR). A fourth innovation of our database is the transformation of texts through OCR to achieve one

single format. This can allow a more detailed information extraction of those treaties through text mining, deriving high-quality information from text, not limited by coding.

5.3. Outcome of the Research

This project had two main objectives: (i) to create a new database of IIAs that complement existing endeavours, including texts of bilateral investment treaties (BITs), and other treaties with investment provisions (TIPs) in one single language (English) and format and easily accessible by creating a machine-readable structured database; and (ii) to analyse how IIAs are negotiated, structured, and diffuse over time, as well as whether they influence investment flows.

As detailed both in the Final Research Plan (27.08.2015) and in the Detailed Timeline (27.10.2015), the Project was divided in three phases: i) Data Generation and Extraction (originally planned between October 2015 – August 2016); ii) Research Tasks (originally scheduled for August 2016 – July 2017); and iii) Dissemination Tasks (planned for July 2017 – October 2017).

The first general observation for the execution of the Project is that the planned activities started not before the end of November 2015, as the SNIS funding was only received on 24 November 2015. Therefore, some activities were initiated with almost one month of delay in comparison to what was originally planned. The second general observation is that the project had a considerable delay in the data generation phase, which partially affected the research output.

Below is a description of the main activities that have been undertaken in each phase:

5.4. Phase 1: Data Generation and Extraction

a) Treaty retrieval

With the objective of having a systematic database on international investment agreements (IIAs), the first task of this phase of the Project was dedicated to identifying a complete list of IIAs that were missing from existing databases, and to obtain their texts. This task was performed mostly by the Bern-based team (Dr Rodrigo Polanco, Mr Dmitriy Skugarevskiy, and research assistants Mr Valentino Desilvestro and Ms Azernoosh Bazrafkan), with the help of research assistant Ms Lamiya Baz from Kuwait, and some volunteers from different countries around the world (Ms Faith Tigere from South Africa, Ms Ngan Nguyen from Vietnam, and Mr Sebastian Espinosa from Ecuador).

In a first step, it was necessary to agree on a definition of what we understood as an IIA. We have worked with following definition “An international investment agreement is a treaty concluded under international law between two or more states or economies, 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 (BITs), investment chapters of preferential trade agreements (PTAs) and Regional Investment Agreements (RIAs) that have been traditionally included in existing databases. It also includes some agreements that have been largely excluded from them, like Friendship, Commerce and Navigation Agreements (FCNs) and parts of PTAs outside of their investment chapter (relevant parts on services, financial services and taxation) and side or additional agreements on investment. This definition further excludes from its scope other types of

agreements that have been included in existing databases (largely by UNCTAD) such as 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.

In a second step, in order to obtain a consolidated list of IIAs and its texts, a standardized research procedure ('Hunting IIAs') was followed that included:

- i. Searching in existing public and private databases, both digital and physical.¹
- ii. Searching in governmental websites – for that purpose a comprehensive list of 139 different websites of ministries, agencies and public institutions was developed, after extensive inspection of websites.
- iii. Formally contacting governments and international organizations (e.g. World Bank, OECD) in order to obtain a copy of the official text of agreement.

The original task was envisaged to take place between October 2015 and January 2016, but some unexpected challenges affected the timing of concluding this task. The compilation of a first batch of around 2315 treaties was finalized on 16 February 2016 followed by successive updates, the latest on 22 October 2016. Several reasons account for this gradual process:

- i. Investment treaties are a moving target. Since the submission of the pre-proposal until the date of this report, at least 52 new IIAs have been reported.
- ii. Existing databases – particularly UNCTAD's International Investment Agreements – have updated their content, addressing the gap. For example, by early 2015, UNCTAD reported 3489 IIAs with 817 missing texts (386 of them in force). After an important update on 22 September 2017, UNCTAD reports 3608 IIAs, but only 3134 available texts (491 texts missing, including 455 BITs (165 in force) and 19 TIPs (7 in force)).²
- iii. Some governments were reluctant to share information, and either denied access to the texts, or never answered our requests.
- iv. Some treaties were wrongly indexed in existing databases as IIAs.
 - The section 'treaties with investment provisions (TIPs)' in the UNCTAD database largely refers to FTAs. However, not all FTAs that are signed have negotiated a separate investment chapter, such as the 2013 Chile-Thailand FTA,³ the 2012 EU-Central America Association Agreement.⁴ The 2002

¹ The main databases used to populate our database are the following: UNCTAD's International Investment Agreements [<http://investmentpolicyhub.unctad.org/IIA>], United Nations Treaty Collection [<https://treaties.un.org/>], Kluwer Arbitration "BITs" [<http://www.kluwerarbitration.com/CommonUI/BITS-countries.aspx>], Oxford's "Investment Claims" [<http://oxia.oup.com/#>], Investor-Law Guide – ISLG [<http://www.investorstatelawguide.com/>], Investment Treaty Arbitration – ITA [<http://www.italaw.com/investment-treaties>], and ICSID's "Investment Laws of the World" (only in printed loose-leaf).

² These changes have important effects on our analysis. For example, until the latest update of UNCTAD database, the text of 18-19 % of BITs in force signed by Finland and Sweden are unavailable. Today All Finnish BITs are available on UNCTAD and only 7% of Swedish BITs are missing.

³ <http://www.sice.oas.org/Trade/CHL_THA_Final/CHL_THA_FTA_Full_Version_PDF_e.pdf> and <<http://investmentpolicyhub.unctad.org/IIA/CountryOtherIias/41#iiaInnerMenu>> both accessed on 6 March 2017.

⁴ <<http://trade.ec.europa.eu/doclib/press/index.cfm?id=689>> and <<http://investmentpolicyhub.unctad.org/IIA/CountryGroupingTreaties/17#iiaInnerMenu>> both accessed on 6 March 2017.

agreement between Czech Republic and Slovakia is also not a BIT, but an agreement to terminate the 1992 BIT according to EU law.⁵ There is also a number of BITs that do not contain any investment related provisions, but are named as such. Those include mere trade agreements,⁶ Double Taxation Agreements (DTAs),⁷ economic cooperation agreements,⁸ and Memoranda of Understanding.⁹

- Some agreements were listed as two separate treaties, whereas in reality these had identical texts. This was the case for treaties signed by countries that split up in various states (e.g. Serbia and Montenegro and other countries that formed Yugoslavia).
- Amendments or additional protocols of IIAs were considered as separate agreements. In our dataset, these documents were manually added to the original treaty that has been modified.
- For many non-English-speaking countries, the IIAs were not easily found on the official government website while the Double Taxation Agreements were

⁵ <<http://www.mfcr.cz/cs/legislativa/dohody-o-podpore-a-ochrane-investic/prehled-platnych-dohod-o-podpore-a-ochra>> and <http://europa.eu/rapid/press-release_IP-15-5198_en.htm> both accessed on 6 March 2017.

⁶ The 2010 Guyana-Kuwait BIT is not a BIT but a trade agreement <<http://investmentpolicyhub.unctad.org/IIA/country/112/treaty/1850>> and <http://www.minfor.gov.gy/docs/foreign_trade/Guyana-Kuwait.pdf> accessed on 6 March 2017.

⁷ The 2001 Kuwait-Syria BIT is not a BIT, but a double taxation agreement <<http://www.mof.gov.kw/TaxationFAQ/AgreementViewer.aspx?pdfpath=Syria//394d9a14-ad28-448f-b63d-30a4986aec42.pdf>>, the 2003 Mauritania-Qatar BIT is not a BIT, but a double taxation agreement <<http://www.almeezan.qa/AgreementsPage.aspx?id=1787&language=ar>>, the 2003 Qatar-Syria BIT is not a BIT, but a double taxation agreement <<http://www.almeezan.qa/AgreementsPage.aspx?id=1250&language=ar>>, the 2000 Qatar-Yemen BIT is not a BIT, but a double taxation agreement <<http://www.almeezan.qa/AgreementsPage.aspx?id=1255&language=ar>>, the 1998 Qatar-Sudan BIT is not a BIT, but a double taxation agreement <<http://www.almeezan.qa/AgreementsPage.aspx?id=1520&language=ar>> all accessed on 6 March 2017.

⁸ The 2007 Angola-Spain BIT is not a BIT, but a financial cooperation agreement <<http://investmentpolicyhub.unctad.org/IIA/country/5/treaty/3544>> and <<http://www.icex.es/icex/es/navegacion-principal/todos-nuestros-servicios/informacion-de-mercados/paises/navegacion-principal/el-pais/relaciones-bilaterales/index.html?idPais=AO#5>>, the 2006 Djibouti-Italy BIT is not a BIT, but an agreement for the conversion of debt in public investment for development <<http://investmentpolicyhub.unctad.org/IIA/country/58/treaty/1299>> and <http://itra.esteri.it/Ricerca_Documenti/wfrmRicerca_Documenti.aspx>, the 2002 Malta-Italy BIT is not a BIT, but an economic cooperation agreement <<http://investmentpolicyhub.unctad.org/Download/TreatyFile/1694>>, the 2000 Belarus-Libya BIT is not a BIT, but a cooperation agreement in education, science and culture <<http://naviny.org/2000/11/01/by58094.htm>>,

the 1996 Armenia-Turkmenistan is not a BIT, but two agreements: one on cooperation in the field of public health and the other on exchange of legal information <<http://www.parliament.am/library/erkoxm/1996-1998/7.pdf>> and <<http://www.parliament.am/library/erkoxm/1996-1998/8.pdf>>, the 1994 Kazakhstan-Mongolia BIT is not a BIT, but a labor contract <<http://adilet.zan.kz/kaz/docs/P1400001422>>, the 1973 Kuwait-Tunisia BIT is not a BIT but two agreements: one on cultural, technical and information exchange and the other on economic cooperation <[http://www.diplomatie.gov.tn/index.php?id=399&no_cache=1&L=0&tx_wdtreaties_pi1\[country\]=KWT&tx_wdtreaties_pi1\[action\]=listView&tx_wdtreaties_pi1\[pointer\]=1&tx_wdtreaties_pi1\[mode\]=1](http://www.diplomatie.gov.tn/index.php?id=399&no_cache=1&L=0&tx_wdtreaties_pi1[country]=KWT&tx_wdtreaties_pi1[action]=listView&tx_wdtreaties_pi1[pointer]=1&tx_wdtreaties_pi1[mode]=1)> all accessed on 6 March 2017.

⁹ The 2000 Indonesia-Venezuela BIT is not a BIT, but a Memorandum of Understanding for the implementation of bilateral consultations <http://treaty.kemlu.go.id/uploads-pub/1904_VEN-2000-0005%20.pdf> accessed on 6 March 2017.

fully accessible and available on the website of Ministry of Finance and/or on the website of the designated Investment Promotion Agency.¹⁰ Nevertheless, there were also countries that we did not presume to have all their IIAs available, such as Romania¹¹, Moldova¹², Georgia¹³, Turkey¹⁴, unlike many Arab countries: United Arab Emirates¹⁵ and Algeria,¹⁶ Philippines¹⁷ and many Latin American countries.¹⁸ In contrast, another group of countries lived up to our expectations and has published all their IIAs, such as China,¹⁹ Malaysia,²⁰ and Singapore.²¹

- When visiting the official government websites and the Investment Promotion Agency websites of the various countries, we could find the full text of the negotiated Double Taxation Agreements, while on the same websites we could only find a list of signed IIAs.²² For example, the Moroccan Ministry of Finance provides a list of it signed IIAs but does not refer to the full text of

¹⁰ For example, Seychelles Investment Promotion Agency refers to BITs but has only Double Taxation Agreements available on its website. See <<http://www.sib.gov.sc/index.php/info-centre/downloads/dta-ippas-bilateral-treaties>> accessed on 6 March 2017.

¹¹ <<http://www.mae.ro>> accessed on 6 March 2017.

¹² <<http://lex.justice.md>> accessed on 6 March 2017.

¹³ <<http://www.mfa.gov.ge/MainNav/ForeignPolicy/InternationalLegalAgreements.aspx>> accessed on 6 March 2017.

¹⁴

<http://www.ekonomi.gov.tr/portal/faces/oracle/webcenter/portalapp/pages/content/docListViewer.jspx?folder=/Contribution%20Folders/web/Yat%20C4%20B1r%20C4%20B1m/Yurtd%20C4%20B1%20C5%20F%20C4%20B1na%20Yat%20C4%20B1r%20C4%20B1m/01.Uluslararası%20C4%20B1%20Yat%20C4%20B1r%20C4%20B1m%20Anla%20C5%20Fmalar%20C4%20B1%20%2028YKTK%29/02.Y%20C3%20BCr%20C3%20BCr%20C3%20BCKteki%20Anla%20C5%20Fmalar/&parentPage=yatirim&_afWindowMode=0&_afWindowId=null#!%40%40%3F_afWindowId%3Dnull%26_afWindowId%3D571973888236149%26folder%3D%252FContribution%2BFolders%252Fweb%252FYat%25C4%25B1r%25C4%25B1m%252FYurtd%25C4%25B1%25C5%259F%25C4%25B1na%25C4%25B1r%25C4%25B1m%252F01.Uluslararası%25C4%25B1%25C4%25B1r%25C4%25B1m%25C4%25B1na%25C5%259Fmalar%25C4%25B1%25C4%25B1m%2528YKTK%2529%252F02.Y%25C3%25BCr%25C3%25BCr%25C3%25BCKtek%25C4%25B1na%25C5%259Fmalar%252F%26parentPage%3DYatirim%26_afWindowMode%3D0%26_adf.ctrl-state%3Da9wrmuoh_e_60> accessed on 6 March 2017.

¹⁵

<<http://www.economy.gov.ae/english/Ministry/MinistrySectors/ForeignTradeSector/Documents/Forms/AllItems.aspx>> accessed on 6 March 2017.

¹⁶ <<http://www.invest.caci.dz/index.php?page=accord-bilateral-d-encouragement-et-de-protection-reciproques-des-investissements>> accessed on 6 March 2017.

¹⁷

<http://61.28.185.135/treaty/index.php?r=treaty%2Fsearch&Treaty%2Btitle%2D=&Treaty%2Bcategory_id%2D=7&Treaty%2Bdate_of_signature%2D=&Treaty%2Bdate_of_entry_into_force%2D=&Treaty%2Bcontracting_party_id%2D=&yt0=Search> accessed on 6 March 2017.

¹⁸ <http://www.sice.oas.org/Investment/bitindex_e.asp> accessed on 6 March 2017.

¹⁹ <<http://english.mofcom.gov.cn/article/bilateralchanges/201309/20130900300306.shtml>> accessed on 6 March 2017.

²⁰ <<http://www.miti.gov.my/index.php/pages/view/771>> accessed on 6 March 2017.

²¹ <<https://www.mti.gov.sg/MTIInsights/Pages/IGAs.aspx>> accessed on 6 March 2017.

²² For example, Cuba <<https://cubatravelcorp.wordpress.com/2016/01/03/acuerdos-de-promocion-y-proteccion-reciproca-de-inversiones-appri/>>, Mozambique <<http://www.cpi.co.mz/index.php/en/investment-guarantees>>, Burkina Faso <<http://www.apexb.bf/les-accords-bilateraux-regissant-les-exportations>>, and Yemen <<http://investinyemen.org/content.php?c=24&langid=1&pageid=3>> accessed on 6 March 2017.

those agreements.²³ Other examples are IPAs that have published their domestic legislation with regard to the treatment of foreign investment (Investment Code) but remain silent on the signed IIAs, such as Cambodia and Mozambique.²⁴ This implies that host countries (developed and developing alike) perceive the Double Taxation Agreements as instruments to attract foreign investment, but not the IIAs.

- Finally, some errors in the metadata of IIAs (date of signature, date of entry into force), also made difficult to locate the treaties. On some occasions, the signature date of the IIA is not correctly mentioned in the UNCTAD database, such as the 2009 France-Syria BIT,²⁵ the 2003 Slovenia-Uzbekistan BIT,²⁶ the 1998 Belgium-Luxemburg Economic Union (BLEU)-Yemen BIT,²⁷ the 1995 Bangladesh-Pakistan BIT,²⁸ and the 1979 Germany-Nigeria BIT.²⁹ In case of Canada-Slovakia BIT of 1997, it seems that this agreement never existed. There is only one previous BIT between Canada and Slovakia prior to the 2010 BIT which dates from 1990 that was signed with Czech and Slovak Republic.³⁰

As mentioned in the Detailed Timeline of the Project, it is important to note that we started to collect the relevant parts of other treaties related to foreign investment – such as commitments undertaken under the WTO’s General Agreement on Trade in Services (GATS) “mode 3” (commercial presence) and non-discrimination provisions in double taxation agreements (DTAs), to be used in the research activities of the second phase of the Project. But both because of the delay of the data production, and due to the number of difficulties that we faced to integrate different texts in one single, referred above, we could not achieve one of our objectives of integrating our database with relevant parts of GATS “mode 3” commitments and DTAs. However, it is important to recall that, as it was clarified in the detailed final timeline of the project, the data collected from existing databases on trade in services and tax

²³ <<http://www.finances.gov.ma/Docs/2013/Accord%20de%20non%20double%20imp%20arabe.pdf>> accessed on 6 March 2017.

²⁴ Cambodia <<http://www.cambodiainvestment.gov.kh>>, Mauritius <<http://www.investmauritius.com/downloads/ipa.aspx>> both accessed on 6 March 2017.

²⁵ <<http://investmentpolicyhub.unctad.org/IIA/country/72/treaty/1615>> and <http://basedoc.diplomatie.gouv.fr/exl-php/cadcgp.php?CMD=CHERCHE&QUERY=1&MODELE=vues/mae_internet_traites/home.html&VUE=mae_internet_traites&NOM=cadic_anonyme&FROM_LOGIN=1> both accessed on 14 June 2017.

²⁶ The only BIT between Slovenia and Uzbekistan dates back to 1999 <https://www.slov-lex.sk/static/pdf/1999/101/ZZ_1999_101_19990518.pdf> accessed on 6 March 2017.

²⁷ The text of the agreement was negotiated in 1998 but was not signed until 2000 <<https://www.senate.be/www/?Mlval=/publications/viewPub.html&COLL=S&LEG=2&NR=936&VOLGNR=1&LANG=fr>> accessed on 6 March 2017.

²⁸ The agreement was signed in 1988 and not in 1995 <<http://investmentpolicyhub.unctad.org/IIA/country/16/treaty/380>> and <<http://www.clcbd.org/document/741.html>> accessed on 6 March 2017.

²⁹ Germany did not sign an agreement with Nigeria on that date but with Oman <<http://investmentpolicyhub.unctad.org/IIA/country/78/treaty/1728>> and <<http://dipbt.bundestag.de/doc/btd/10/021/1002121.pdf>> accessed on 6 March 2017.

³⁰ UNCTAD <<http://investmentpolicyhub.unctad.org/IIA/country/35/treaty/800>> , See also UNCTC <<https://treaties.un.org/doc/Publication/UNTS/Volume%202469/v2469.pdf>> and Slovakian Ministry of Finance <<http://www.finance.gov.sk/Default.aspx?CatID=4597>> accessed on 6 March 2017.

commitments (WTO database for GATS commitments; OECD database for DTAs), was only for research purposes, and it was not intended that our new IIA database would duplicate work.

b) Digitization of treaties

After obtaining the official texts of the IIAs, the next step in our Project was to convert them to a single digital format. We decided to use the Extensible Markup Language (XML) format considering that it defines a set of rules for encoding documents that can be both human-readable and machine-readable, and allows electronic annotations in the text. Considering that IIAs include several articles, and those articles usually are made up of several paragraphs, treaty texts are split up to paragraph level, with the purpose of capturing the highest level of detail possible in the subsequent annotation of the legal texts. The texts were therefore augmented with mark-ups for (1) layout, (2) text structures, and (3) linguistic information.

This task was performed mostly by the University of Zurich-Institute of Computational Linguistics (ICL) based team (Professor Martin Volk, Dr Kyoko Sugisaki) with the help of Bern-based Mr Dmitriy Skugarevskiy and Dr Wolfgang Alschner.

The challenge to transform the texts in a single format arose especially for the diverse type of sources in which this data was found. Around 1623 treaties which were split at article level were already converted in XML format from a previous related research project run by Dr Wolfgang Alschner and Mr Dmitriy Skugarevskiy (<http://mappinginvestmenttreaties.com/#> 'SNIS Geneva XML'). This original dataset was expanded to include around 2100 IIAs split at paragraph level.

The rest of the treaties were found in different formats (PDF, Microsoft Word, and HTML). Among them, pdf documents were the most prevalent, but with different quality across texts. We roughly divided them in "Type A" (originally in paper form and scanned or photographed, usually of lower quality) and "Type B" (originally word documents converted into PDFs, usually of higher quality) documents. For example, from the first batch of treaties 534 treaties were classified as Type A and 390 as Type B. We have so far processed 1026 international investment agreements in HTML, Microsoft Word, and PDF format in more than 30 languages. A few more are in the pipeline.

To bring these heterogeneous agreement texts into one format and one language, we have processed PDF documents through state-of-the-art optical character recognition (OCR) software (Abbyy Recognition Server), translated multi-lingual texts into English, and converted unstructured text data into structured data.

The original timeframe for this task was considered to be carried out between December 2015 and March 2016, but it also took longer for following reasons:

- i. Delays in treaty retrieval, as explained in the previous section.
- ii. Conversion problems especially for those treaties where the text was found only in lower-quality pdf (Type A).
- iii. Some multi-lingual documents that include more than one language at the level of sentences, columns, and pages in a document presented distinctive problems of

- language recognition. To solve them, an n-gram³¹ based language identification software was used together with manual examination and correction of the texts.
- iv. The conversion of treaties from more than 20 languages posed challenges because the software does not automatically identify these languages, and also certain languages have posed more problems for language recognition (e.g. Arabic, Chinese).
 - v. Some documents had to be manually corrected after being converted to XML, as the process of conversion is not error-free.

Even though our source data comes in a variety of formats, the layout of various texts possesses a strong commonality (they contain text blocks and paragraphs), that facilitated the mark-up process. Layout mark-ups were enhanced to also reflect document structure, as IIAs comprise preface (such as title page and table of contents), preamble, text body (i.e. article and paragraph), conclusion (i.e. signatures), and sometimes attachments (e.g. annexes). Hence, we divided each document into these five text zones and annotated the zones with XML mark-up. For the segmentation of these five text zones, we compiled a set of corresponding, typical linguistic feature patterns. In the text body and attachments, we automatically structured texts at the level of articles and paragraphs. This segmentation was again based on the surface patterns of enumeration structures.

We further enriched this XML mark-up with linguistic information for English. For this purpose, we used a tool of natural language processing. Texts were automatically segmented into sentences, and then the sentences were tokenized and augmented with part-of-speech tags, named entities identifiers, and dependency grammar structures. This linguistic information allows us to extract further information, for example, metadata such as the date and place of signature of the treaties.

Yet, this process was not exempt of problems. The most important one was the lack of adequate delineation of XML files at article level. The only solution for this problem was manual correction, that involved an unforeseen number of working hours. For example, the last correction of IIAs, involved the cleaning of 151 PTAs with investment chapters in good shape. While the number of treaties might appear small, in fact those treaties comprise 18266 articles of 67997 in our data in total, i.e. 26.9%. Therefore, this is not an insignificant or small-scale correction.

Other problems that affected this output, were typos in file structure (something that is prone to happen in long texts), some problems with ABBYY conversion routines and lack of support in certain languages (notably in bad quality Arabic texts), and lack of adequate development in treaty structure detection routines (i.e. detecting which chunks of texts constitute articles, which — sections).

c) Language Availability

When BITs are available, in 81% of cases they are available in only one language (see Table 2 last row). Interestingly, approximately 90% of BITs signed among Asian countries are more likely to be available in more than one language. When a treaty is available in only one language, it can either be the language of either country, a common language, or a language of a third party.

³¹ An n-gram is a contiguous sequence of n items from a given sequence of text or speech. The items can be phonemes, syllables, letters, words or base pairs according to the application. The n-grams typically are collected from a text or speech corpus.

Treaties signed between an Asian country and a country in the North are most likely to be available in a third language, while treaties signed between an Asian and an African country are the most likely to be available only in the language of one country.

Table 2: Language availability, by region

Region	Available	1Lang	1Lang %	3rd %	Common %	Either %
North	464	366	79	58	2	43
Africa	118	109	92	15	28	57
Africa-S.America/Carr	19	16	84	25	12	62
Africa-Asia	170	154	91	42	23	35
Africa-North	382	299	78	29	14	58
S.America/Carr	80	70	88	1	89	10
S.America/Carr-Asia	89	76	85	50	1	49
S.America/Carr-North	316	233	74	29	13	59
Asia	297	257	87	62	5	33
Asia-North	733	578	79	58	3	40
Sum	2,668	2,158	81	46	11	44

Available indicates the total number of treaties signed, at least one text available.
 1Lang indicates the count of treaties available in only one language, 1Lang % the percentage. The remaining columns break down the composition of 1Lang into:
 - % of treaties available in a third language (3rd);
 - % of treaties available in a common (official) language (Common);
 - % of treaties available in either official language (Either).

Table 3 reports some more detailed statistics for treaties that are available in only one language, restricting the analysis to cases in which the language of the treaty is one of the language of the two parties (common language cases are excluded). For instance, there are 233 BITs signed by Asian countries with a country in the North that are available only in one of the official languages of the two parties (excluding common language). In 71% of those cases, the text is available only in the language of the country in the North. Similar calculations can be done by using income level classifications. 77% of BITs signed by a low-income country with a middle-income country are available only in the language of the low income country.

Table 3: BITs available in language of either party

Party1-Party2	N (only either lang)	Lang Party1	Lang Party2	Lang Party1 %	Lang Party2 %
Africa-S.America/Carr	10	2	8	20	80
Africa-Asia	54	38	16	70.37	29.63
Africa-North	172	76	96	44.19	55.81
S.America/Carr	7	6	1	85.71	14.29
S.America/Carr-Asia	37	33	4	89.19	10.81
S.America/Carr-North	137	59	78	43.07	56.93
Asia-North	233	68	165	29.18	70.82
High-Low	64	27	37	42.19	57.81
High-Middle	480	315	165	65.62	34.38
Low-Middle	62	48	14	77.42	22.58

The column N displays the number of treaties that are available in only one of the languages of the two parties. The remaining indicates in which language of the two parties the treaty is available.

The most common languages are English, French, Spanish, Arabic, Russian and German (See Table 4). 77% of treaties available in English in the UNCTAD database are available only in English, and 65% of those use English as a lingua franca.

Table 4: Most common languages

Language	N	N %	N _{,only}	N _{,only} %	3rd %	Common %	Either %
En	1,850	58.49	1,424	76.97	64.96	4.42	30.62
Fr	465	14.70	280	60.22	17.86	10.36	71.79
Es	278	8.79	186	66.91	0	42.47	57.53
Ar	144	4.55	59	40.97	0	89.83	10.17
Ru	141	4.46	73	51.77	2.74	2.74	94.52
De	108	3.41	35	32.41	0	0	100

N and N% indicate the number and percentage of treaties available in the relevant language.

N_{,only} and N_{,only}% indicate the number and percentage of treaties available only in the relevant language. The remaining columns break down the composition of N_{,only}:

- % of treaties available as a third language (3rd);
- % of treaties available as a common (official) language (Common);
- % of treaties available as either official language (Either).

d) Machine translation of treaties

About 64% of the IIAs found in our Project are written officially in English or have a complimentary translation into that language. Other important languages are French (around 10% of the treaties) and Spanish (around 7% of the treaties), followed by Russian and Arabic (each one representing around 3% of the treaties), and German (around 2% of the treaties). All other languages represent less than 1% of the universe of treaties.

In this task, we committed to translate all non-English treaties to that language using tailor-made software for the IIAs in French and Spanish (that constitute the largest part of non-English treaties) and using general machine translation engines (Google Translator) for the few IIAs that are exclusively in other languages (Russian, Arabic, German, and other languages like Italian and Portuguese, among others), with manual corrections by team members, research assistants or volunteers fluent in those languages. This task was performed mostly by the University of Zurich-based team (Professor Martin Volk, Dr Annette Rios, and Ms Laura Mascarell) with the help of Bern-based Dr Rodrigo Polanco, Dr Wolfgang Alschner, and Mr Valentino Desilvestro, and of Ms Lamiya Baz from Kuwait.

First, two machine translations were built for treaties in French and Spanish ('MOSES'), using Joint Research Centre (JRC) texts from the EU, and 'parallel texts' (originally in Spanish/English and French/English in the source IIAs) provided by the Bern-based team. Overall, 2000 sentences were used for tuning. The outcome of this machine translation provided high accuracy in short texts and common languages. However, compared with Google Translator (GT), in the first exercise GT performed better, especially in technical texts, as probably Google incorporates information that feeds the system every time GT is used.

On a second stage MOSES was further improved with data from UNCorpus (11.6 million sentences) and extracted from SNIS Geneva XML data (90k sentences). MOSES translation was then compared with GT 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.

The outcome of this new machine translation effort provided increased accuracy in comparison with the previous exercise, but Google Translator (GT) still has better BLEU scores, except in very specific treaties. GT and MOSES have different strengths and are

insofar complementary, and either machine translation pipeline has deficiencies. MOSES is good at recognizing legal imperatives and legal registers and GT is better at correctly translating abbreviations, negate sentences and is better at recognizing and fixing spelling mistakes.

University of Zurich-based team is currently working in fine-tuning MOSES, using a trilingual dictionary of state-of-the-art terms used in IIAs elaborated by the Bern-based team, and correcting some minor problems detected about abbreviations, spellings and negations.

The original timeframe for this task was considered to take place between January and June 2016, but it also became an on-going one, and it was finally completed in October 2017. Several reasons influenced this outcome:

- i. Delays in treaty retrieval, as explained in previous sections
- ii. Delays and text problems in the digitization of treaties, as explained in previous sections
- iii. The unexpected good performance of a general machine translation engine (Google Translator), that has pushed the team to refine tailor-made machine translation to even higher levels.

e) Treaty Categorisation (Preparation of Data)

In this task, we commit to prepare the data to enable researchers to analyse the data using different types of software. For that purpose, we examined and compared treaty texts, identifying recurrent expressions in order to create rules to enable the computational markings or 'labelling' of treaty articles and sections. The outcome of this process must be validated with the different research teams of the second phase of the project, to assess the quality and the limitations of the data acquired.

The idea is that once all the treaties are in a single format (XML) and language (English), they can be automatically 'tagged' or labelled with conceptual categories, through intrinsic text (e.g. article header when they are available in the treaty text) or extrinsic text (e.g. keywords coding). For that purpose, the Bern-based team has developed a 'dictionary' with different categories that is currently under revision by the whole team and external collaborators from the German Development Institute and from the Pluricourts project from the University of Oslo. This task has been performed mostly by the Bern-based team (Prof. Manfred Elsig, Dr Wolfgang Alschner, Dr Rodrigo Polanco, Mr Dmitriy Skugarevskiy, and research assistant Mr Valentino Desilvestro), with inputs from University of Zurich-based team (Professor Martin Volk and Dr Kyoko Sugisaki) and of Professor Peter Egger from ETH (Chair of Applied Economics).

As the final 'tagging' will require important precision (as it should take place at a paragraph level), we have been exploring how to do improve the automated labelling. For that purpose, we tested different software (e.g. 'Oxygen'), to do manual annotations on treaty texts corpora in XML, allowing labelling from a pre-defined list. This allowed us to get a precise benchmark on how automated annotation should look. For the final categorisation we used an intelligent scope-limiting algorithm, thereby not categorising articles not related to

investment in the treaties, something that is particularly important in PTAs with investment treaties.

The original timeframe for this task was considered to take place between February and August 2016, but it also became an on-going one and it was finally completed in October 2017. Several reasons influenced this outcome:

- a) Delays in treaty retrieval, as explained in the previous sections
- b) Delays and text problems in the digitization of treaties, as explained in previous sections
- c) Delays in machine translation, as explained in previous sections.
- d) The complexities in the construction of the Dictionary, as it requires manual reading of several treaties (extrinsic texts) and the manual control of errors derived from the automated labelling based on article headers (intrinsic texts).

5.5. Phase 2: Research Tasks

The second phase of the Project considered four research tasks that considered initially of an analytical phase of five months (August – December 2016), and the proper writing of the papers over six months (January – June 2017). As explained before, due to the delays of the data production of the project, some of the plan research could not be completed or undertaken. Given the high motivation of team members and the current societal relevance for research on investment treaties several research activities were advanced using existing data. Some of the papers have already been accepted for publication. Team members are regularly invited to present ongoing work at various conferences.

5.5.1. Outcome of Planned Research

The original research tasks envisaged for the project included:

5.5.1.1. Information Extraction from Treaties

First, we engaged in information extraction from text based on simple regular expressions, performing a linguistic analysis of the texts, consisting of lemmatization, Part-of-Speech tagging and dependency parsing. For each sentence, this gave us the syntactic functions such as subject and object, as well as temporal and causal relations. The parser recognized the subject of the sentence, as well as its syntactic objects. It also determined the core statement of the sentence which establishes and labels the relation between the subject and the object.

Automatic syntactic parsing is not perfect, but stills help to identify the roles of the various actors and objects in the treaties. One challenge of this project was the adaptation of the English parser to the legal language of IIAs. In particular we needed to handle the many and complex coordinated constructions and the resulting ambiguities of their modifiers. Legal texts are very structured, yet their structure is not easily extractable. Our team developed a method that parses the treaties for us identifying chapters, articles, paragraphs and annexes. This task was completely achieved, as described in the previous sections of this report.

5.5.1.2. Network Analysis and Economic Effects

In our research, we have engaged with two broad classes of network analysis. The first relates to the topology of the space of treaties. As with the work of Baccini, Dür & Elsig [2015] on measuring the depth of PTAs, the goal here is to provide a more nuanced analysis of the content of IIAs.

This task was only partially achieved in our research, as it is described in the outputs section below, especially with respect of the topology of countries revealed by investment treaties, identifying common patterns, rule-makers or rule-takers, and the current effective gap for a multilateral investment treaty. In the economic and network analysis, we aimed to examine the topology of the space of investment treaties and its effect of different components of IIAs on FDI flows. Now that the data has been collected our researchers can try to identify provisions of IIAs that could have an impact on FDI flows, examining similarities and different models of IIAs.

5.5.1.3. The Adoption, Copying and Evolution of Treaty Language

EDIT also allows us to investigate the way in which IIAs are negotiated, how its language is adopted and evolves, and how much of that language is unique to a particular treaty compared to pre-existing agreements.

We aimed to quantify the amount of text in a given IIA that is drawn from previous treaties, comparing pairs of agreements and producing a percentage match for the amount of text shared by any two treaties, focusing on specific clauses in IIAs, as we suspect that the origins and mechanisms of diffusion vary significantly across provisions. Finally, we aimed to create a summary indicator of how much content in a particular IIA is unique versus taken from other existing IIAs.

With these techniques, we planned to assess how much boilerplate language there is in IIAs, identifying overlapping patterns across treaties to find places where the exact same language is evoked, and how treaty text diffuses over time and space (from agreement-to-agreement or country-to-country), tracking changes in language over time. This task was only partially achieved in our research, as it is described in the outputs section below

5.5.1.4. Integrating Investment Commitments

EDIT serve as the basis to estimate the overall level of international commitments for any state with respect to foreign investment and foreign investors. It also allow us to study specific commitments, such as granting pre-establishment rights that are usually subject to country-specific exceptions. Another example is the content of most-favoured nation (MFN) provisions that allow the import of more favourable investor rights (either substantive or procedural) from other investment treaties. According to MFN clauses, a state that grants a more favourable treatment to another state is obliged to extend to third states the treatment accorded between the granting and the beneficiary state in the new treaty. MFN clauses therefore level the differences found in the standard of protection offered by different IIAs. Restrictions on the operation of MFN clauses will be considered as a limit to this analysis, including those that explicitly confine the clause itself (for example, not including dispute settlement rights), and imply limitations (such as indirect restrictions based on the scope of application of the IIA, or of the MFN clause) [Schill, 2014].

Our project aimed to identify the “invisible” common investment provisions, encompassing a country’s overall level of commitments on foreign investment undertaken with respect to third states under IIAs and related treaties. This would allow some measurement of the distance between the current bilateral system and an eventual multilateral investment agreement. It will also help to compare the positions of the negotiating countries, before and after concluding a specific IIA, based on previous IIAs or model treaties.

This task was only partially achieved in our research, as it is described in the outputs section below.

5.5.1.5. Expanding the Toolkit of Lawyers Through Computational Legal Analysis

The toolkit of international lawyers is geared towards solving “small data” problems. Drafting an investment treaty, writing an expert opinion or litigating an investment claim requires attention to detail. Yet, faced with thousands of IIAs, international lawyers also increasingly face problems of a “big data” nature. As litigators, lawyers have to be able to identify similarities and differences between agreements to effectively reason through precedent or find more favourable treatment importable through MFN clauses. As negotiators or advisors to policymakers, lawyers have to ensure coherence in a country’s treaty network or identify common denominators from hundreds of treaties to prepare the grounds for multilateral law-making. The traditional toolkit of lawyers is ill equipped to address these “big data” challenges. Big data problems instead require big data solutions.

EDIT expanded the lawyers’ toolkit by surveying computational text analysis methods suitable for application in “big data” legal analysis. Computational text analysis developed as part of natural language processing. Given the different structure of legal language, the same tools cannot be blindly transposed to the legal field. Instead, we will take stock of existing supervised and unsupervised machine learning tools in order to match relevant legal research questions to available research methods identifying a novel toolkit for computational legal analysis.

Using EDIT, we aimed to provide practical guidance to legal scholars and lawyers within and beyond the field of investment law. This research yields new insights into the current state and trajectory of investment treaty law, and related research using EDIT database will assist litigators in effectively comparing the similarity of underlying treaties and help negotiators in finding a common ground in regional or global rule-making efforts. EDIT has also produced a more general toolkit that can be expanded and applied to other fields of law in order to solve big data legal questions elsewhere.

This task was almost completely achieved in our research, as it has been described in the section introducing the content of the EDIT database, as well as described in the outputs section below, including the use of EDIT data in existing websites like mappinginvestmenttreaties.com.

5.5.2. Current Research Output

Below we list some of the key research papers that have been published up to now, and provide for a short abstract:

- a) “Who wins BITs negotiations?” (Dr Wolfgang Alschner, Dr Eddy Bekkers & Dmitriy Skougarevskiy). The goal of the paper is to explain negotiation outcomes on international investment agreements (IIAs). More in particular we are interested in the question who “wins” negotiations on Bilateral Investment Treaties (BITs). To do so we consider the similarity of a BIT signed by negotiating country i with partner country j with the IIAs signed thus far by country i , S_{ij} . Similarity is an indicator of whether country i has been able to negotiate a BIT in line with its preferences and has thus “won” the negotiations. As an alternative, we study the difference in similarity

between IIAs signed thus far by country i and IIAs signed thus far by country j , DS_{ij} . This working paper was presented in the abovementioned workshop that took place in Bern at the WTI on December 9, 2016.

- b) “Missing Investment Treaties” (Dr Rodrigo Polanco, Azernoosh Bazrafkan & Valentino Desilvestro). The goal of this study is to explain the reasons behind the lack of public information on IIAs, particularly the absence of texts of these agreements, even in cases where these treaties are in force. Our research also addresses the implications of this lacuna in the existing literature on effects of IIAs in foreign direct investment flows. How one can claim that these treaties have any effect (positive or negative), if their text (or sometimes the fact that they exist) is unknown? This working paper was presented in the abovementioned workshop that took place in Bern at the WTI on December 9, 2016, and it was submitted for consideration to be published in the Journal of International Economic Law (JIEL).
- c) “Building a Corpus of Multi-lingual and Multi-format International Investment Agreements” (Dr Kyoko Sugisaki, Professor Martin Volk, Dr Rodrigo Polanco, Dr Wolfgang Alschner; & Mr Dmitriy Skougarevskiy). In this paper, we present an on-going research project whose aim is to develop a new database of international investment agreements that complements existing endeavours. In particular, this paper describes our efforts to build a standardized corpus of multi-lingual and multi-format agreement texts in order to enable researchers in the fields of international law and economics systematically investigate investment treaties. This paper was presented at Jurix 2016 Conference that will take place in Nice, France on 15 & 16 December 2016.
- d) “International Investment Agreements and G-20 countries” (Dr Rodrigo Polanco, Ms Azernoosh Bazrafkan, Mr Valentino Desilvestro and Ms Lamiya Baz). The objective of this paper is to map the international investment agreements concluded by G-20 countries, with the goal to examine whether there are common patterns among this group of countries regarding investment protection. For that purpose, the most important provisions on investment protection were examined in all treaties and across all G-20 countries. These provisions include the scope and definition of investment and investors; the existence of establishment or pre-establishment commitments; standards of treatment (national treatment, most favoured nation, full protection and security, fair and equitable treatment) and their limitation; main standards of protection, like expropriation (and its standard of compensation, and the protection against indirect expropriation), transfers (and the presence of exceptions based on problems on the balance of payments or alike), and umbrella clauses. Finally we examined the presence of investor-state arbitration provisions and whether they were limited by MFN clauses. The dataset used in this research was the one generated by EDIT complemented with UNCTAD’s database of investment treaties and mapping. This paper was presented at the World Economic Forum Dialogue “2016 G20 Investment Principles in the broader trade and investment agenda - Advancing the

G20 outcomes” that took place in Geneva on 7 November 2016, and was published as part of the E-15 working papers.³²

- e) “Mapping the Universe of International Investment Agreements” (Dr Wolfgang Alschner & Mr Dmitriy Skougarevskiy). Traditional means of content analysis are ill-equipped to deal with the vast universe of international investment agreements (IIAs). In this article, we propose a novel approach to efficiently investigate over 2100 IIAs and their 24,000 articles in unprecedented detail by treating treaty text as data. Our suggested metric yields new and surprising insights about the IIA universe at four different levels. First, at the global level, we use our approach to investigate the effect of asymmetries on negotiation outcomes finding that developed countries tend to be the IIA system’s rule-makers, while developing countries tend to be its rule-takers. Second, on the country level, our method can trace consistency and legal innovation in national treaty networks uncovering hitherto unknown investment policy changes such as the Finnish shift to a pre-establishment template in 1999. Third, on the inter-treaty level, our metric can detect investment policy diffusion highlighting that Israel, for instance, copied its bilateral investment treaty (BIT) language from British investment agreements. Finally, on the individual treaty level, our approach enables us to assess the novelty of newly concluded agreements, like the Trans-Pacific Partnership, by relating them to prior practice. Our metric thus provides researchers, practitioners and policy-makers with a powerful novel tool to analyse the IIA universe. This article was published at the *Journal of International Economic Law*, Vol. 19, No. 3, 2016, pp. 561-588.
- f) “Rule-takers or Rule-makers? A New Look at African Bilateral Investment Treaty Practice” (Dr Wolfgang Alschner & Mr Dmitriy Skougarevskiy). Who are the rule-takers and rule-makers in the African BIT universe? Using computational measures of textual similarity this paper provides a nuanced empirical answer to this question. First, we find that African states tend to be rule-takers in North-South relations, yet enjoy greater agency in negotiations of South-South BITs. Only a few African countries, however, use their greater say in intra-African negotiations to include public policy exceptions in BITs. Indeed, recent North-South BITs contain more policy space than their Southern counterparts in Africa. Second, rule-makers and rule-takers also exist within the African South-South BIT landscape. Yet, in contrast to North-South relations, negotiation outcomes seem to be shaped more by expert knowledge than by power asymmetries. Powerful states like Egypt fail to dominate negotiations, while small-island-state Mauritius with its strategic investment policy agenda succeeds in setting the terms of investment agreements. This paper thus provides a more nuanced view of the African treaty landscape, points to new areas of research and highlights the importance of technical expertise in achieving coherent treaty networks. This article was published at the *Transnational Dispute Management*, Vol. 13, Issue No 4, 2016.

³² <http://e15initiative.org/themes/investment-policy/>

- g) “A Requiem for the TPP. Something new, something old, and something borrowed?” (Dr Rodrigo Polanco Lazo and Mr Sebastián Gómez Fiedler). In this article, the authors have compared the on the core disciplines of TPP with the texts of the previous treaties concluded between TPP signatory states. Information from the EDIT database was used to compared the PTAs with investment chapters concluded by TPP members with the TPP investment chapter. This article will be published in the forthcoming number of the Melbourne Journal of International Law (Vol. 18 No. 2, December 2017)
- h) A Master Thesis by Xi Rao, titled “Automatic Labelling of Articles in International Investment Agreements Using Semi-Supervised Learning and Word Embeddings”, was co-supervised by Profs. Martin Volk and Peter Egger (submitted 3 July 2017, at the Faculty of Arts and Social Sciences of the University of Zurich, for the degree of Master of Arts UZH). In this master thesis, the author attempted to automatically assign titles to untitled international investment treaty articles using semi-supervised learning. Various titles have been assigned to similar texts due to the variability of negotiating partners, languages, traditions, etc. Hence, in order to have a condensed representation of various article titles, the author firstly cluster 34,524 titled articles into ten topics by expanding word and document semantics with embeddings. The author then used these ten classes as the labels in our classification task where titles are assigned to 10,074 untitled articles. The classification task is performed with supervised classification ers (k-nearest neighbours (KNN), support vector machine (SVM), multi-layer perceptron (MLP), stochastic gradient descent (SGD), convolutional neural network (CNN)) and partially supervised k-means clustering. Expert annotations of 100 untitled articles are used as the gold standard in our evaluation set. K-means clustering with the retrained word embeddings tailored to our corpus has brought about an increase of 30% in accuracy compared to a simple CNN classier, which has scored the highest amongst all supervised classifiers. The comparison between these two machine learning paradigms (supervised and semi-supervised learning) leads to the conclusion that word embeddings can effectively expand the semantic features for words and documents, which allows us to perform accurate categorization of texts from closely related sub-fields of one research area, for instance, to categorize the ten topics in the study of IIAs.³³
- i) “Shaping Investment Rules: The Imprint of International Organisations on Bilateral Investment Treaties (Wolfgang Alschner, Manfred Elsig, Dmitriy Skougarevskiy, and Simon Wuethrich). This paper was presented at the Annual Meeting of the American Political Science Association Meeting in August 2017. It measures through textual similarity which IOs through model treaties and recommendations have had the greatest impact on the evolution of BITs.

We will inform the SNIS after the submission of the final report about the release of any publications, articles and other outputs linked to project, and to submit one hard copy and, if

³³ https://www.cl.uzh.ch/dam/jcr:30fcec3e-c1cd-47da-930f-fc5295036152/Masterarbeit_XRao_FS2017.pdf

possible, an electronic version, or otherwise, the exact references of the output so that it can be published on the SNIS website.

5.6. Phase 3: Dissemination

The third phase of the Project considered a mid-term workshop and final conference at the end of the project (September 2017) to help disseminate the project's research findings and present the newly created database.

- a) A workshop to assess the first year of the SNIS project took place at the WTI on 9 December 2016. The workshop was divided into three parts: a stock-taking of the first year was followed by the presentation of research papers on the negotiation of investment agreements and the availability of their texts. The final part was a brainstorming on a roadmap for the future of the project. The event was attended by researchers at the WTI, the University of Zurich (Institute of Computational Linguistics), the ETH Zurich (Chair of Applied Economics), the German Development Institute (Deutsches Institut für Entwicklungspolitik - DIE) and from the University of Oslo's PluriCourts project.³⁴
- b) An investment conference took place at the World Trade Institute (WTI) under the title 'Is a Multilateral Investment Treaty Needed?', on 19 June 2017, implementing one of the main lines of research of the SNIS project (the convergence of investment treaties and the possibility of a multilateral investment agreement). Besides the support of SNIS, the conference was sponsored by the World Economic Forum (WEF) having Transnational Dispute Management (TDM) as the media partner. The conference, attended by 85 people, focused on the many important questions emerging from the negotiation of a possible multilateral framework agreement on investment, and reflected in the dramatic changes taking place in the investment treaty landscape. The details of the conference are mentioned in the working paper.

The conference was divided into four sessions: Background, the EU Proposal, The Relationship with Existing Treaties, and Future Steps. Ali Dadkhar of Ciuriak Consulting began the first session, chaired by Manfred Elsig, by addressing the question of whether investment agreements promote foreign investment. He was followed by WTI senior researcher Rodrigo Polanco, who talked about the hunt for missing bilateral investment treaties (BITs) and the reasons why they cannot all be found. Malebakeng Forere of the University of Witwatersrand spoke about the move away from BITs and why there needs to be a multilateral investment treaty, while Robert Basedow of the European University Institute talked about lessons to be learned from the failed Multilateral Agreement on Investment (MAI) negotiations in the late 1990s.

Session II, chaired by Ana Novik of the Organisation for Economic Co-operation and Development (OECD), looked at the EU proposal for a multilateral investment court, with Colin Brown and Maria Luisa Andrisani presenting the rationale for this, and its possible design features. Elsa Sardinha of the National University of Singapore

³⁴ <https://www.wti.org/institute/news/398/snis-project-workshop-takes-stock-of-first-year/>

discussed the possibility of using the two-tier investment tribunal system found in the EU-Canada CETA agreement and the EU-Vietnam FTA as the basis for a multilateral effort. She was followed by Anna De Luca of the University of Bocconi and José Manuel Alvarez of Universidad Externado, Colombia.

The relationship with existing treaties was the subject of the third session, chaired by Krista Nadakavukaren Schefer of the University of Basel, with speakers Iryna de Meyer from the Energy Charter Secretariat, N. Jansen Calamita from the National University of Singapore, Martin Björklund from the University of Helsinki and Markus Wagner of the University of Warwick.

The final session, chaired by Nikos Lavranos, European Federation for Investment Law and Arbitration (EFILA), looked at future steps and included presentations by Facundo Calvo of the University of Barcelona, Lise Johnson of the Columbia Center on Sustainable Investment, Junqing Chao, University of Hong Kong, and Axel Berger and Wan-Hsin Liu of the German Development Institute and Kiel Institute for the World Economy.

The event closed with a discussion and policy roundtable involving Nathalie Bernasconi-Osterwalder (IISD), Ricardo Meléndez Ortíz (ICTSD), Corinne Montineri (UNCITRAL), Pierre Sauvé (World Bank), and Elisabeth Tuerk (UNCTAD).

The papers presented at the conference will be considered for publication at the European International Arbitration Review.

- c) The general aspects of the EDIT database were presented by Prof. Manfred Elsig at the 2017 WTO Public Forum, in the Working Session No. 64, entitled “Rules Behind the Headlines: using data to navigate the unseen landscape of international economic law”, organized by the Centre for Trade and Economic Integration, Graduate Institute of International and Development Studies. This session discussed how big data and new technologies can help stakeholders to better manage the growing complexity of international economic law and to more fully reap the benefits of the rules behind the headlines. Research output based on several newly-established databases on WTO disputes, FTAs, and BITs respectively will be presented. These databases pull together efforts of lawyers, economists, political scientists and computer scientists, using state-of-the-art “text as data” tools to reduce international economic law’s complexity and make trade rules work behind the headlines, but for everyone.

5.7. Collaboration between Partners

Collaboration between partners functioning particularly well during the first year of the project:

- a) The WTI Bern-based team met regularly with the University of Zurich-based team of the Institute of Computational Linguistics with seven meetings during the year, alternating meeting places between Bern and Zurich (14 September 2015, 9 November 2015, 21 March 2016, 2 May 2016, 21 June 2016, 30 August 2016 and 17 October 2016).

- b) The WTI Bern-based team met two times with Professor Peter Egger from ETH at the WTI (3 October 2016 and 26 October 2016).
- c) The WTI Bern-based team met with Professor Doug Nelson (Tulane University), who provided feedback on the working papers described in Section 1.1.2 and of the overall functioning of the project.
- d) The WTI Bern-based team met with Professor Doug Nelson (Tulane University), at Galbino Summer School, in Florence, Italy (18-20 July, 2016), who provided feedback on the working paper described in Section 1.1.2a) and of the overall functioning of the project.
- e) The WTI Bern-based team met with Mr Roberto Echandi (World Bank), at the WTI on 15 April 2016 and 21-22 July 2016, who provided feedback on the working paper described in Section 1.1.2b) and of the overall functioning of the project.
- f) Exchanges with other Project participants such as Professor Todd Allee (University of Maryland) and Professor Andrew Newcombe (University of Victoria) have taken place via e-mail and Skype sessions.

During 2017 the coordination activities with partners were mostly done online, with the exceptions of the dissemination activities described in the preceding section and the Galbino Summer School that took place in Italy (July 2017), where several team members attended.

5.8. Difficulties Encountered

The scientific and practical difficulties encountered in Phase 1 and 2 of the Project have been previously described in the preceding sections of this paper.

Several personal difficulties appear over the duration of the Project:

- a) Ms Magdalena Plamada (ICL) decided to stop her work within the Project after completing her maternity leave by mid-2016. She was replaced by Dr Annette Ríos and Ms Laura Mascarell both at ICL, under the same budget originally approved.
- b) Mr Dmitriy Skugarevskiy had problems in obtaining his visa that impede him to get a contract, a process that took from November 2015 until April 2016. For that reason, part of his salary was advanced using a consultancy agreement, in order to help him to meet his financial needs and to make sure that the project did not get further derailed. This decision did not affect the budget originally approved. Furthermore, Mr Skugarevskiy, faced some personal problems that impacted in his ability to deliver his part of the data in due time.
- c) Dr Rodrigo Polanco's salary was miscalculated in the original budgeted, as the money allocated could only cover 40% of his salary plus social charges and not 50% as it was considered in the budget originally approved. In order to keep the budget in line with the one approved by SNIS, the difference of 10% was covered by the WTI.
- d) Dr Kyoko Sugisaki could not continue her engagement with the project after February 2017, and her work could not be finished due to a delay in the processing of the last batch of data. This led to a lacuna in the data generation process that was only filed in July 2017, by her replacement with Dr Annette Rios. It is important to highlight that these extra costs were directly assumed by the WTI, outside of the SNIS budget.
- e) In May 2017, Dr Eddy Bekkers one of the co-coordinators of the project, left the WTI and started to work full time at the World Trade Organization (WTO).

It is also important to note that the World Bank advanced its commitment to provide research assistants and engaged Ms Azernoosh Bazrafkan and Ms Lamiya Baz to support the Project during 2016 (for 20'000 CHF), keeping the remainder of the budget originally planned for 2017 (for 40'000 CHF) for the future development of the dedicated and 'user-friendly' website where the database developed during the SNIS Project will be migrated to during the first semester of 2018.

5.9. Bibliography

Allee, T. and Elsig, M. (2014), "Are the Contents of International Treaties Boilerplate? Evidence from Preferential Trade Agreements", Annual Meeting of the American Political Science Association Washington, D.C., August 28-31

Alschner, W. and Skougarevskiy, D. (2015) "Consistency and Legal Innovation in the BIT Universe", Stanford Public Law and Legal Theory Working Paper Series, Research Paper No. 2595288, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2595288

Baccini, L. et al (2015). "Imitation and Innovation in International Governance: The Diffusion of Trade Agreement Design," in Trade Cooperation: The Purpose, Design and Effects of Preferential Trade Agreements," A. Dür and M. Elsig (eds.), Cambridge University Press.

Baccini, L., Dür, A. and Elsig, M. (2015), "Design of Trade Agreements (DESTA) Database", www.designoftradeagreements.org.

Baldwin, R. (2014), "Multilateralising 21st Century Regionalism", OECD, February, 2014, <http://www.oecd.org/tad/events/OECD-gft-2014-multilateralising-21st-century-regionalism-baldwin-paper.pdf>.

Behn, D., Fauchald, O.K. and Langford, M. (2014), "Legitimacy of International Investment Tribunals: Towards a Grounded, Actionable and Empirical Informed Approach". PluriCourts workshop on the legitimacy of international courts and tribunals, Oslo, November 2014.

Berger, A. (2013), "Do We Really Need a Multilateral Investment Agreement?", *DIE Briefing Paper*, Vol. 9/2013, German Development Institute.

Bergera A., et al. (2010), "Do Trade and Investment Agreements Lead to More FDI? Accounting for Key Provisions Inside the Black Box", WTO Economic Research and Statistics Division.

Börzel, T. and Risse, T. (2012), "From Europeanisation to Diffusion: Introduction," *West European Politics* 35(1):1-19.

Brown, C. (2013), "Commentaries on Selected Model Investment Treaties", Oxford University Press.

Chaisse, J. and Bellak, C. (2015), "Navigating the Expanding Universe of International Rules on Foreign Investment - Creation and Use of a Critical Index" *Journal of International Economic Law*, vol. 18(1) pp. 79-115.

Chaisse, J. and Bellak, C. (2015), “Coding the Expanding Universe of International Treaties on Foreign Investment-- The BITSel Index”, <http://www.cuhk.edu.hk/law/proj/BITSel/>.

Cooper, A., Heine, J. and Thakur, R. (2013). *The Oxford Handbook of Modern Diplomacy*. Oxford: Oxford University Press.

Dolzer, R. and Stevens, M. (1995), “Bilateral Investment Treaties”, Martinus Nijhoff Publishers.

Dür, A., Baccini L., and Elsig, M. (2013), “The Design of International Trade Agreements: Introducing a New Database.” *Review of International Organizations*: 1-23.

Ebling et al. (2011), “Digging for Names in the Mountains: Combined Person Name Recognition and Reference Resolution for German Alpine Texts”, in *Proceedings of The 5th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics*, Poznan.

Egger, P. and Merlo, V. (2012). "BITs Bite: An Anatomy of the Impact of Bilateral Investment Treaties on Multinational Firms," *Scandinavian Journal of Economics*, vol. 114(4), pp. 1240-1266.

Egger, P. and Pfaffermayr, M. (2004) "Foreign Direct Investment and European Integration in the 1990s," *The World Economy*, vol. 27(1), pp. 99-110, 01.

Elkins, T. et al (2006), “Competing for Capital: The Diffusion of Bilateral Investment Treaties: 1960-2000”, *International Organization* 60, Fall 2006, pp. 811–846.

Elkins, T. and Simmons, B. (2005), “On Waves, Clusters, and Diffusion: A Conceptual Framework”, *Annals of the American Academy of Political and Social Science* 598, pp. 33–51.

Francois, J. et al. (2013), “Reducing Transatlantic Barriers to Trade and Investment. An Economic Assessment”. http://trade.ec.europa.eu/doclib/docs/2013/march/tradoc_150737.pdf

Gaukrodger, D. (2014), “Investment Treaties and Shareholder Claims for Reflective Loss: Insights from Advanced Systems of Corporate Law”, *OECD Working Papers on International Investment*, 2014/02, OECD Publishing.

Hallward-Driemaier, M. (2003), “Do Bilateral Investment Treaties Attract FDI? Only a Bit and They Could Bite”, *World Bank Policy Research Paper*, WPS 3121 (Washington, D.C.: World Bank).

Höfler, S. and Sugisaki, K. (2014), “Constructing and exploiting an automatically annotated resource of legislative texts”, in: *Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, 26 May 2014 - 31 May 2014, pp. 175-180.

Investment Arbitration Reporter – IA Reporter (2015), <http://www.iareporter.com/>.

Jandhyala, S., Henisz, W., and Mansfield, E. (2011). “Three Waves of BIT s: The Global Diffusion of Foreign Investment Policy.” *Journal of Conflict Resolution* 55(6), pp. 1047–73.

Kerner, A. and Lawrence J. (2012), “What's the Risk? Bilateral Investment Treaties, Political Risk and Fixed Capital Accumulation”, *British Journal of Political Science* 44(1), pp. 107-121.

Manger, C. and Peinhardt, C. (2014), “ Learning and Diffusion in International Investment Agreements”, *The Political Economy of International Organizations (PEIO) conference at Princeton University*, January 16-18, 2014, http://www.uni-heidelberg.de/md/awi/peio/manger_peinhardt_26.08.2013.pdf.

Molinuevo, M. (2012), “Protecting Investment In Services. Investor State Arbitration Versus WTO Dispute Settlement”, *Alphen aan den Rijn, Kluwer Law International*.

Neumayer, E. and Spess, L. (2005), “Do Bilateral Investment Treaties Increase Foreign Direct Investment to Developing Countries?” *World Development* 33 (10), pp. 1567-1585.

Organisation for Economic Co-operation and Development (OECD) - Global Forum on Transparency and Exchange of Information for Tax Purposes (2015), “Exchange of Tax Information Portal”, <http://eoi-tax.org/#default>.

Park, W. (2001) *Arbitration And The Fisc: NAFTA's “Tax Veto”*, *Chicago Journal of International Law* 2, Spring 2001, pp. 231-241.

Park, W. (2009), *Arbitrability and Tax*, in L Mistelis and S Brekoulakis (eds), *Arbitrability: International & Comparative Perspectives*, Netherlands: Kluwer Law International.

Pluricourts (2014), *Investment Treaty Arbitration Database*, <http://www.jus.uio.no/pluricourts/english/topics/investment/research-projects/database.html>.

Pohl, J. (2013), “Temporal Validity of International Investment Agreements: A Large Sample Survey of Treaty Provisions”, *OECD Working Papers on International Investment*, 2013/04, OECD Publishing.

Pohl, J., K. Mashigo and A. Nohen (2012), “Dispute Settlement Provisions in International Investment Agreements: A Large Sample Survey”, *OECD Working Papers on International Investment*, 2012/02, OECD Publishing.

Polanco Lazo, R. (2015), “The Return of the Home State to Investor-State Disputes. Bringing Back Diplomatic Protection?”, *PhD Thesis, University of Bern, The Graduate School of Economic Globalisation and Integration, World Trade Institute*, 2015.

Poulsen, L. (2010), “The Importance of BITs for Foreign Direct Investment and Political Risk Insurance: Revisiting the Evidence”, http://works.bepress.com/laug_poulsen/4.

Poulsen L., et al. (2013) “Costs and Benefits of an EU-USA Investment Protection Treaty” https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/260380/bis-13-1284-costs-and-benefits-of-an-eu-usa-investment-protection-treaty.pdf.

Putnam, R. (1988), “Diplomacy and Domestic Politics: The Logic of Two-Level Games”, *International Organization* 42(3), pp. 427-460.

Radin, M. (2013). *Boilerplate: The Fine Print, Vanishing Rights, and the Rule of Law*. Princeton: Princeton University Press.

Salacuse, J. and Sullivan, N. (2005), “Do BITs Really Work? An Evaluation of Bilateral Investment Treaties and Their Grand Bargain”, *Harvard International Law Journal* 46(1), pp. 67-130.

Schill, S. (2014), *The Multilateralization of International Investment Law*, Cambridge University Press, January 2014.

Schneider, G. (2008), “Hybrid Long-Distance Functional Dependency Parsing”, PhD Thesis, University of Zurich, Institute of Computational Linguistics, 2008.

Sennrich, R. et. al (2009), “A New Hybrid Dependency Parser”, in: *Proceedings of GSCL-Conference*. Potsdam, 2009.

Sennrich, R. (2012), “Perplexity Minimization for Translation Model Domain Adaptation in Statistical Machine Translation”, in *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, Avignon, France, 2012.

United Nations Conference on Trade and Development – UNCTAD (2007), “Bilateral Investment Treaties 1995-2006: Trends in Investment Rulemaking”, United Nations.

United Nations Conference on Trade and Development – UNCTAD (2013-2014), “IIA University Mapping Project”, <http://investmentpolicyhub.unctad.org/Upload/Documents/UNCTAD%20IIA%20MAPPING%20PROJECT%202013-2014.pdf>.

United Nations Conference on Trade and Development – (2015), “Recent Trends In IIAs And ISDS”, *1 IIA Issues Note*, http://unctad.org/en/PublicationsLibrary/webdiaepcb2015d1_en.pdf.

Volk, M. et al. (2011), “Strategies for Reducing and Correcting OCR Errors”, in Sporleder et al. (eds.): *Language Technology for Cultural Heritage: Selected Papers from the LaTeCH Workshop Series*. Springer-Verlag, Berlin, 2011, pp. 3-22.

World Economic Forum (2013), “Foreign Direct Investment as a Key Driver for Trade, Growth and Prosperity: The Case for a Multilateral Agreement on Investment”, <http://www.weforum.org/reports/foreign-direct-investment-key-driver-trade-growth-and-prosperity-case-multilateral-agreement>.

World Trade Organization - WTO (2015), Schedules of commitments and lists of Article II exemptions, https://www.wto.org/english/tratop_e/serv_e/serv_commitments_e.htm.

World Trade Organization and the World Bank (2015), “I-TIP Services”, [http://i-tip.wto.org/services/\(S\(hzkhkwlviyvch1vdpxtsnii\)\)/default.aspx](http://i-tip.wto.org/services/(S(hzkhkwlviyvch1vdpxtsnii))/default.aspx).

Yackee, J., (2010) “Do Bilateral Investment Treaties Promote Foreign Direct Investment? Some Hints from Alternative Evidence”, University of Wisconsin Legal Studies Research Paper No. 1114.