

2. Executive Summary

2.1. Research Plan

2.1.1. Problem Statement

Investment is an important driver of the global economy. It interacts closely with trade and impacts other fields like development, environment, labour mobility and taxation. The global investment regime is governed through a fragmented network of intergovernmental treaties known as international investment agreements (IIAs). At the core of this regime are around 2,950 bilateral investment treaties (BITs). The first generation of these agreements were negotiated at the end of the 1950's and were mostly concluded between developed and developing countries. This pattern changed during the 1990's when the number of BITs grew exponentially, and developing countries and economies in transition signed BITs among themselves. In addition, investment chapters started to be included in certain preferential trade agreements (PTAs) following the example of NAFTA Chapter 11, which can be considered the first investment treaty between two developed countries, Canada and the US.

IIAs include definitions and commitments related to the very notion of foreign investment and foreign investors, and standards of investment protection, including national and most-favoured nation treatment, fair and equitable treatment, full protection and security, protection against expropriation, rights of transfer and subrogation. Notably, most IIAs allow foreign investors to challenge governmental actions outside local courts, using investor-state dispute settlement (ISDS) provisions to trigger binding investor-State arbitrations. ISDS cases have proliferated in the past 15 years, with the overall number of known treaty-based arbitrations reaching 817 by the end of October 2017. In addition, ISDS has become politicized not only because of the immense sums involved in the claims, but also for concerns related to its legitimacy and impartiality. We witness opposition both in developed and developing countries, especially in the context of the negotiations of the "mega-regional" agreements, such as the Comprehensive Economic and Trade Agreement (CETA), the Trans-Pacific Partnership (TPP) and the Transatlantic Trade and Investment Partnership (TTIP).

No other field of international law is characterized by such a high number of intergovernmental treaties. Switzerland is one of the top three countries in terms of the number of BITs (114), following Germany (133) and China (129). Also, while some model-type agreements have been highlighted, there is substantial variation of obligations across treaties and over time. Additionally, the broader investment regime encompasses obligations derived from double taxation agreements (DTAs) and commitments on trade in services undertaken under the General Agreement on Trade in Services (GATS). The existence and effects of overlapping commitments have been largely ignored. Today, the available databases of IIAs lack comprehensives which hamper the study of the design and effects of these agreements. Existing collections of texts, while useful, provide partial information, are only

available in the original language negotiated by parties, and exist in formats that do not lend themselves easily for textual analyses.

2.1.2. General Objective

This project has had two main objectives:

First and most importantly the project aims to create a new database of investment treaties that complements existing endeavours, including integrated texts of BITs, and the relevant parts of PTAs, DTAs, and GATS commitments in one single language and format. For that purpose, besides the extraction of data from the treaties, advanced analytical techniques were used, including the digitization and machine translation of treaties, modelling and development of algorithms for the purpose of information extraction, with a 'research friendly' software interface. This new database is aimed to assist scholars and practitioners alike to study the similarities and differences of legal texts over time and across treaties.

Second, with the information obtained from the database, the project attempts to push the research frontier by applying advanced statistical methods and computational linguistics digital tools to contribute to on-going debates in international law, international relations and international economics. We aim to analyse the degree of "boilerplating" that is occurring and how treaty texts diffuse. In addition, we aim to map the structure of the network of investment treaties and measure the distance between the current bilateral system and an eventual multilateral agreement. New treaty design measures is also considered to be examined under outcomes such as investment flows. Finally, it is envisaged to examine in more details the outcomes of investment treaty negotiations and studying the learning and diffusion process throughout the whole network of agreements.

2.1.3. Research Questions

Several research questions have guided our work:

- Is there a common language of investment treaties? How has it evolved in terms of temporal or spatial diffusion?
- Who are the main creators and key vectors of that language? What is the impact of level of development, power, competition, emulation on treaties?
- Can different models be identified on the texts of IIAs? Which have most elements in common and which countries are the ultimate sources of a certain model?
- Can we identify which specific components of IIAs are most important in the promotion FDI flows?
- How can we link the patterns of the network with investment flow effects?
- Can we define the depth of IIAs in a similar way as the depth of FTAs and has the depth of IIAs an effect on FDI flows and affiliate sales?
- What is the level of overall commitment undertaken by a state with respect to foreign investors and foreign investment? Are there overlaps or points of conflict between BITs, PTAs, GATS commitments and DTAs with respect to investment? If so, is this reflected in the text of the IIAs?
- What is the level of concordance between IIAs? How far we are from a multilateral investment agreement?

2.2. Research Methods and Organisation of the Project

The project aimed to compile in a single database all available IIAs, in a format that would allow a computational analysis of its text, for purposes of document retrieval, text recognition, translation, text mining and document analysis. The research using the newly compiled data relies on a combination of quantitative and qualitative methods. The choice of mixed methods methodology intends to address key research questions on design, diffusion and effects of investment treaties. Different methods were used in the phases of data generation, information extraction and data analysis and in each specific research task, as described below.

The project was divided in two main stages: (1) data generation and (2) data analysis. The data generation included the following three tasks: text retrieval, text recognition and text translation. Data analysis included information extraction of investment treaties using advanced techniques of computational linguistics and network analyses.

2.2.1. Data Generation

2.2.1.1. Treaty Retrieval

For the purpose of organizing the information on the IIAs, we initially relied on UNCTAD's IIAs database metadata, including parties, date of signature and coming into force, as well as termination date if applicable. This was complemented with additional information collected by the research team from government websites, with requests under domestic laws. However, as several countries do not make public the text of negotiated agreements until the treaties have undergone domestic approval, the database aim was to retrieve the texts of all IIAs currently in force, complemented by those agreements that have been signed, inasmuch as their texts are publicly available.

2.2.1.2. Digitization of Treaties

A large number of treaties are only available as scan images but not as electronic text. We will employ off-the-shelf OCR (Optical Character Recognition) software (e.g. *Abbyy FineReader*) to recognize the text and save it in digital form. However, OCR works with a certain rate of misrecognized words depending on the quality of the scan image, the font of the original text, and the number of words unknown to the OCR system dictionary.

Recognition errors fall into two basic categories: misrecognition to a real-word or to a non-word of the language. A real-word error can only be detected by analysing the context and is thus a hard challenge for automatic methods. We tackled this in our project using context-aware real-word error detectors along the idea of grammar checking and domain-specific vocabulary.

In contrast, the detection of a non-word error is a relatively simple spell-checking task. However, the automatic correction of such errors is still not trivial. Methods for auto-correction of these errors have been developed by computing the most similar word (in terms of character distance) from the corpus. But, as error-free electronic text is the desired goal of any digitization involving legal texts, then manual checking was necessary.

2.2.1.3. Machine Translation of Treaties

Although English can be considered the lingua franca of IIAs, a large number of these treaties are written in a language other than English (like French, Spanish, Arabic, Russian and

German), without many translations into English made publicly available. In order to automatically analyse the treaties and to extract information in a uniform manner, we defined that it was needed to have all treaties in the same language, and that the most common language (English) was the most efficient choice to normalize the translation.

To automatically translate the treaties into English, we used a mixed method: on the one hand, freely available machine translation (MT) systems (*Google and Bing Translator*), and on the other hand, a special purpose MT for the languages that had a sufficient number of treaties that justify the effort of building a dedicated software (French and Spanish). The assumption was that MT systems are general purpose translation engines, meant to translate all kinds of text as good as possible, and therefore there were less reliable. We assumed that special purpose MT systems would result in higher translation quality when tuned for a given text type – in this case, IIAs.

For the other IIAs that are exclusively in other languages, general MT engines were used (especially Google Translator), with manual post editing as a control mechanism from people proficient in the respective language and with knowledge on IIAs. Some treaties required manual translation, as the OCR process did not produce intelligible texts that were susceptible of being translated.

2.2.2. Data Analysis - Specific Research Tasks

As part of the second phase of the project, different team members were supposed to work with the data and apply state-of-the-art methods to answer a number of specific research questions. This work was divided in five areas:

- a) Information extraction from treaties
- b) Network analysis and economic effects
- c) The adoption, copying and evolution of treaty language
- d) Integrating investment commitments
- e) Expanding the toolkit of lawyers through computational legal analysis.

2.3. Schedule

The original schedule included the following milestones:

- October 2015 – August 2016: Collection of IIAs, translation to a single language (English), work on computer models and exploratory empirical analyses.
- September 2016 – August 2017: Further empirical analysis, adjustment of computer models and applications for a textual analysis of IIAs. Drafting of working papers and set-up of the database.
- September 2017: Final project workshop (presentation of working papers, external speakers and commentators, policy outreach)
- October 2017: Dissemination of the database on a web page.

For the reasons explained with detail in the working paper, it was not possible to follow this schedule within the planned time frame, as we encountered unexpected delays. In particular, it took longer to finalize the database which affected the writing of papers based on the new data. Group Members have worked on methodological issues and have produced a number of papers that use selected parts of the data.

2.4. Results Obtained and Analysis

The project achieved its main objective: the creation of a more comprehensive database of international investment agreements: The Electronic Database of Investment Treaties (EDIT), provisionally available through following link: https://gitlab.cl.uzh.ch/snis/xml_repository.¹

EDIT includes more available texts than those included in other existing databases, with the texts of all treaties in one single language (English), and in one format ('SNISDocXML') that allows their categorisation (both automatically and based on a keyword dictionary) and manual annotation.

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 closing 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.

It is 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.

2.4.1. Do the Results Correspond to Those Expected?

The main goal of the project was achieved, with the creation of a comprehensive database of IIAs, in one single language and format (EDIT). This will prove important for future research on investment law. However, the successive delays in the generation of the data, consumed large part of the work of the project in this phase, and limited the work that was intended to be carried out during the research phase of the project, as the final version of the database was only available by October 2017. Team members, however, have started several paper projects so the immediate impact will become visible in 2018.

As detailed in the working paper, some of the most important problems in compiling the database, were the difficulties in finding "Missing IIAs", the conversion of bad quality files (especially pdfs of low quality to XML), the lack of uniform article delineation of the treaties, and the categorization of treaties based on categories external to the text for research purposes, especially with respect of "false positives" (parts of the treaty text that could be automatically labelled in more than one category of our research dictionary).

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 other treaties related to foreign investment – General Agreement on Trade in Services (GATS) "mode 3" commitments (commercial presence) and non-discrimination provisions in double taxation agreements (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 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 the work of those existing databases, aiming only to interaction with those existing databases.

Finally, although the database is available on a GitLab repository, its format still does not lend itself for an easy use, although we are already working together with the World Bank to

¹ To access the link is necessary to use the LDAP username: rpolanco, followed by the password: KxWQr5eH

be able to migrate the dataset to a more “friendly-user” website (something that in any case, was not considered in the SNIS budget) during the first semester of 2018.

2.4.2. Information Regarding the Practical Application of Results

Now that EDIT has been completed, the possibilities of practical application of this database can fully be exploited, for all different types of research activities described in this project, particularly with respect to the comparison of the texts of investment treaties, its categorisation according to the different types of labels that are now embedded in the text of the IIAs that are part of the database, that will make easier to search throughout different texts. At the same time, this type of information will facilitate the studies about the diffusion of investment treaty provisions across countries and over time.

In fact, the source information of the EDIT database, is currently being used in a website that is administered by two members of the Project (Dr Wolfgang Alschner and Mr Dmitry Skougarevskiy): Mapping Investment Treaties,² where users can situate and compare texts of investment treaties in the IIA Universe. In the forthcoming website where the current repository of the EDIT database will be migrated, users should be able to further explore and search different categories of provisions across different IIAs, besides getting access to full texts in English, with their respective embedded metadata.

2.4.3. Questions That Merit Further Exploration

Team members have learned a lot from interacting across disciplines. While cooperation was not always easy, the benefits already in collecting the data and preparing it for research usage have been important. One important question that merits further attention is the sustainability of the data collection and the interactive nature of the website. We will explore further how future agreements can be easily integrated, using automated approaches.

2.4.4. Practical and Policy Recommendations

It is too early for the project to make practical or policy recommendation.

2.4.5. Information Regarding Past and Expected Publications and other Activities

The project has received during its short time period already ample attention. Also, team members have been active in organizing workshops, conferences and working jointly on papers.

Conferences and Workshops

- 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

² <http://mappinginvestmenttreaties.com/>

Development Institute (Deutsches Institut für Entwicklungspolitik - DIE) and from the University of Oslo's PluriCourts project.³

- 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, addressing one of the key research questions 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 co-sponsored by the World Economic Forum (WEF) and worked with Transnational Dispute Management (TDM) as a 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 on the important changes taking place in the investment treaty landscape. The details of the conference are mentioned in the working paper.
- 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.

Articles and Papers

- a) "Who wins BITs negotiations?" (Wolfgang Alschner, Eddy Bekkers & Dmitriy Skougarevskiy). 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" (Rodrigo Polanco, Azernoosh Bazrafkan & Valentino Desilvestro). 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" (Kyoko Sugisaki, Martin Volk, Rodrigo Polanco, Wolfgang Alschner; & Dmitriy Skougarevskiy). 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" (Rodrigo Polanco, Azernoosh Bazrafkan, Valentino Desilvestro and Lamiya Baz). 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.⁴

³ <https://www.wti.org/institute/news/398/snis-project-workshop-takes-stock-of-first-year/>

⁴ <http://e15initiative.org/themes/investment-policy/>

- e) “Mapping the Universe of International Investment Agreements” (Wolfgang Alschner & Dmitriy Skougarevskiy). 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” (Wolfgang Alschner & Dmitriy Skougarevskiy). This article was published at the Transnational Dispute Management, Vol. 13, Issue No 4, 2016.
- g) “A Requiem for the TPP. Something new, something old, and something borrowed?” (Rodrigo Polanco Lazo and Sebastián Gómez Fiedler). 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).⁵
- 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.

⁵ https://www.cl.uzh.ch/dam/jcr:30fcec3e-c1cd-47da-930f-fc5295036152/Masterarbeit_XRao_FS2017.pdf