One could think that the universe of international investment agreements (IIAs) has few surprises left to offer. After hundreds of investment tribunals have interpreted its obligations and thousands of scholarly contributions have analyzed its provisions the contours and content of the IIA universe should be well understood. Yet, have you heard about the revision of the Finnish model bilateral investment treaty (BIT) in the late 1990s or Japan’s decision to radically redesign its investment treaty template in 2002? Did you know that the United Kingdom’s treaty network is twice as consistent as that of Egypt or Pakistan? Have you noticed that Israel copied from British treaties to design its own BIT program or that 81 per cent of the Trans-Pacific Partnership’s (TPP) investment chapter is the same as the investment chapter in the United States–Colombia free trade agreement (FTA), concluded ten years before, in 2006?

This is only a selection of the IIA universe’s surprises that we discovered as part of our mappinginvestmenttreaties.com project. In that project, we treat investment treaty texts as data following an approach similar to what is employed in plagiarism detection software. Through our analysis we seek to equip policy-makers, practitioners and researchers with a more...
sophisticated understanding of the IIA universe and better open-access web-based tools to analyze it. In this note we explain the main concepts behind the research, outline our method and showcase the key findings.

**Measuring consistency of treaty language**

Countries tend to strive for consistency in their investment treaty networks. From a host country perspective, consistency facilitates compliance, as only one set of commitments has to be observed rather than a range of varying and potentially conflicting obligations with respect to investors from different partner countries. From a home country perspective, a consistent treaty network offers investors with a predictable baseline of protection wherever they choose to make their investment. Frequently, states create model treaties in order to achieve such consistent treaty networks. These model agreements enshrine what a country considers to be an ideal investment treaty, making deviation from that template in negotiations generally undesirable.

Consistency across treaties can be measured. When two treaties are consistent, for example, because they derive from the same model, they will employ similar or identical language. Although theoretically negotiators are free to choose different words to convey similar ideas, in practice treaty-makers rarely risk varying language merely for stylistic reasons. To ensure consistency and predictability, they instead tend to reproduce established terms and phrases, and only depart from prior formulations where they actually want to express legal differences. Hence, differences in legal language are often associated with differences in legal meaning. As a result, the textual similarity between two treaties is a good indicator of their legal consistency.

We measure the textual similarity between two treaties using the procedure set out in Figure 1. We first split each treaty into its five-character components. Second, we compute the Jaccard distance—a common measure of dissimilarity—between two treaties based on the five-character components that overlap between the pair. While identical treaties will have a Jaccard distance of 0, completely different treaties in which no textual components overlap will have a distance of 1. While the raw Jaccard score between a treaty pair is not very informative, comparing Jaccard scores across treaties yields powerful insights at the system, country and treaty level.

![Figure 1: Illustration of our text-as-data procedure](http://www.iisd.org/itn/wp-content/uploads/2016/05/Alschner-Skugarevskiy-figure1.png)

**System level: Rule-takers and rule-makers**

A first field of application of our metric is the BIT universe as a whole. Prior empirical research suggests that developed countries have often been more successful than developing countries in influencing the outcome of investment treaty negotiations.[1] Our metric now allows us to investigate this claim quantitatively. If developed countries are the system’s rule-makers, there should be more consistency in their treaty networks than in those of developing countries.

To test this claim, we have collected 1628 English-language full texts of BITs and calculated their pairwise Jaccard distances. We visualize the results as a 1628×1628 heat map in which high textual similarity is represented by red cells and low textual similarity by yellow cells. The diagonal line from left to right is red by default, since it compares a treaty with itself.

Two versions of this heat map are displayed: Figure 2(a) is ordered by the wealthier BIT party and Figure 2(b) is ordered by the less wealthy BIT party measured by GDP per capita in the year of treaty signature and other heuristics. While (a) shows red quadrangles that correspond to consistent individual treaty networks primarily of developed countries, (b) shows no similar...
patterns. Our research thus suggests that wealthier countries are the rule-makers, while less wealthy countries are the rule-takers in the BIT universe. The global champion is the United Kingdom, with an average consistency of more than 70 per cent across its vast network of over 100 BITs concluded over a span of 35 years.

**Figure 2: Heat map representation of the BIT universe**

(a) sorted by wealthier party in the treaty pair  
(b) sorted by less wealthy party in the treaty pair

Figure 3 shows the relationship between consistency and economic power by plotting the mean Jaccard similarity (1 minus Jaccard distance) of the treaties a country signed in relation to its GDP. In this figure we do not limit ourselves to English-language treaties, computing per-country textual similarity for French- and Spanish-language treaties as well. Then we average the per-language Jaccard similarities for each country that signed more than five treaties, weighting such similarities by the number of treaties signed in each language.[2] The quadratic trend line in black (with 95 per cent confidence intervals as shaded area) indicates a positive relationship between economic power and consistency across the country’s IIAs, but also allows us to identify two groups of countries: those positioned below the line have treaty networks that are less consistent than it is expected given their level of economic development (for example, Switzerland and Egypt). Conversely, the states above the trend line enjoy surprisingly consistent treaty networks (for example, the United Kingdom and India). Hence, economic power is not the only factor determining whether a country is the maker or taker of IIA rules.

**Figure 3: Consistency across countries’ IIAs and their economic development**

Our research suggests, for instance, that small developing countries with a coherent investment policy and sufficient in-house expertise tend to have more success in ensuring that their preferences prevail in bilateral negotiations. Mauritius (MUS), for example, has successfully established itself as a hub for foreign investment destined for Africa and India. While Mauritius is a rule-taker in IIA negotiations with developed countries, it punches above its weight in negotiations with other developing countries: its treaties signed with India (1998), Ghana (2001) and Egypt (2003) are more similar to Mauritian treaties than to agreements of its more powerful negotiation partners. Hence, a coherent investment strategy and technical expertise can enable even less powerful countries to affect negotiation outcomes.

**Country level: Evolution in national IIA programs**
A second area of application of our metric lies in the analysis of consistency and innovation in national treaty networks. While countries generally strive for consistency in treaty networks, they also want to adjust their investment policy from time to time in response to changing circumstances. For some countries, these changes in investment treaty-making are well documented. The revisions of the U.S. model BIT, particularly the innovations introduced with the 2004 treaty template, have received considerable academic attention. [3] Similarly, scholars have investigated the Canadian BIT program tracing the three-staged transition from a traditional European BIT model to the 1993 integration of North American Free Trade Agreement (NAFTA) elements to the 2004 update of the Canadian BIT template to reflect the country’s experience as respondent in investment treaty arbitration. [4] Using the Jaccard distance, we find these changes in investment treaty-making reflected in the heat map representations of U.S. (http://mappinginvestmenttreaties.com/country?iso=USA) and Canadian (http://mappinginvestmenttreaties.com/country?iso=CAN) treaties on our website.

The fact that our Jaccard distance representation of national BIT networks reveals underlying investment policy patterns can be harnessed to investigate the evolution of national BIT programs that are less well-documented than the above examples. Our analysis, for instance, shows that Japan (http://mappinginvestmenttreaties.com/country?iso=JPN) radically redesigned its BIT program in 2002 and that Finland (http://mappinginvestmenttreaties.com/country?iso=FIN) changed its model BIT in the late 1990s. Understanding patterns in national BIT networks, however, is not only important for researchers, but also for practitioners and policy-makers. Heat map representations of treaties can help identify inconsistencies in national BIT programs. Outlier treaties can become candidates for termination or renegotiation, making a BIT network progressively more consistent. Our interactive website can thus form the starting point for researchers and policy-makers alike to investigate and streamline national investment treaty networks.

**Treaty level: The TPP and other copy-and-paste agreements**

Finally, our metric also helps to situate individual agreements in the wider treaty universe. When a new IIA is concluded, investment lawyers and policy-makers are keen to know whether the new treaty reproduces or departs from existing practice. When the TPP was signed in February 2016, they asked: “How new is the TPP?”

Our metric allows us to provide a nuanced answer. Specifically for the TPP Investment Chapter, we found that its main text overlaps to 81 per cent with the next similar investment treaty—the 2006 United States–Colombia FTA. Moreover, our metric once applied to the article level can reveal where the 19 per cent of divergence from existing practice comes from. While some provisions, like the National Treatment clause, have been copied and pasted into the TPP almost verbatim, other articles, such as the Minimum Standard of Treatment Clause, contain important innovations (http://mappinginvestmenttreaties.com/specials/tpp/). Jaccard distances are thus a powerful tool to quickly identify what is new and what is not in a given agreement.

The same approach can also be used to identify instances of copy and paste more generally. BITs signed by Israel, for instance, look strikingly similar to earlier British agreements, which suggests that Israel was “inspired” by the UK template to design its own BIT program. The Israel–Hungary BIT (1991), for instance, overlaps to 73 per cent with the United Kingdom–Congo BIT (1989). Copy and paste also took place in Cameroon’s BIT practice when it signed almost identical agreements with Guinea, Mali and Mauritania all on the same day—May 18, 2001—during the Third UN Conference on the Least Developed Countries. [5] Our metric can thus be used to trace normative diffusion from one treaty to another.

**Conclusion**

The IIA universe remains full of surprises. The examples presented here are just some of the otherwise hidden consistency trends we came across during our research. Many more of the IIA universe’s secrets still remain to be discovered. We hope that our research and website can assist in that endeavour, allowing scholars to reveal hitherto unknown patterns in treaty-making and assisting policy-makers in streamlining their countries’ investment commitments.

**Authors**

Wolfgang Alschner is a post-doctoral research fellow in international law at the Graduate Institute in Geneva and the World Trade Institute in Bern. Dmitriy Skougarevskiy is a doctoral candidate at the Graduate Institute and a researcher at the Institute for the Rule of Law (EUSPb). The authors gratefully acknowledge the funding support by the SNF Project “Convergence versus Divergence? Text-as-data and Network Analysis of International Economic Law Treaties and Tribunals” and the SNIS Project “Diffusion of International Law: A Textual Analysis of International Investment Agreements.”

**Notes**

[2] To illustrate the approach, consider the case of France. This country has seven English-language BITs, with mean Jaccard distance of 0.59. However, it also has 92 French-language treaties with Jaccard distance of 0.37. Finally, it does not have any Spanish-language treaties. Therefore, the total coherence across 3 languages is \( (0.59 \times 7 + 0.37 \times 92 + 0 \times 0) / (7 + 92 + 0) \approx 0.39 \). So the French treaty network coherence is not 0.59 (if we took only English-language texts, as before) but rather 0.39. The resulting similarity is \( 1 - 0.39 = 0.61 \).


