Working Paper Proceedings

WHO DO SOVEREIGN INVESTORS SAY THEY ARE?

USING MACHINE LEARNING TECHNIQUES TO BUILD A TAXONOMY OF SOVEREIGN INVESTORS

Caroline E. Nowacki, Stanford University, USA

Ashby H. B. Monk, Stanford University, USA

WHO DO SOVEREIGN INVESTORS SAY THEY ARE? USING MACHINE LEARNING TECHNIQUES

TO BUILD A TAXONOMY OF SOVEREIGN INVESTORS

Caroline E. Nowacki[1], Ashby H.B. Monk[2]

ABSTRACT

Sovereign investors (sovereign wealth funds and pension funds) are often referred to as long-term investors. But it is still unclear that these investors pertain to the same category, as they vary on many dimensions. Previous research proposed typologies of sovereign investors based on external criteria such as region or size of assets under management. The recent push for sovereign investors to be more transparent led to the availability of large amounts of "internal" qualitative data. Most sovereign investors now have websites and publish annual reports about their missions, investment strategies, and performance. Through these publications, sovereign investors are playing an active role in defining who they are. We use this large corpus of texts to propose a taxonomy of sovereign investors based on their chosen words. We answer the questions: what themes differentiate sovereign investors, and what external criteria are the most informative to categorize them? We use computer-assisted text analysis and structural topic modeling to analyze a corpus of about 300 annual reports for 36 funds over 10 years to answer these questions. This research contributes to defining the boundaries of the category of sovereign investors, identifies themes that supports several identities, and maps out sovereign investors based on both currently used criteria and themes coming from sovereign investors' own vocabulary. We anticipate this new categorization of investors would help potential partners and organizations assess compatibility in values and identity with these funds. We also wish to present an analytical method new to the field of Engineering Project Organization.

Key words: Sovereign investors; institutional field emergence; organizational identity; computer-assisted text analysis; structural topic modeling.

---

[1] Stanford University, USA, cnowacki@stanford.edu

[2] Stanford University, USA, amonk@stanford.edu

"INFRASTRUCTURE, THE PERFECTLY ALIGNED ASSET CLASS FOR LONG-TERM INVESTORS" This was the title of the 2016 PwC report on Sovereign Investors. Sovereign Funds have consistently increased their asset allocation to infrastructure over the past 10 years. A 2016 Prequin survey of investors indicated that the proportion of Sovereign Wealth Funds (excluding pension funds) investing in infrastructure went from 57% of in 2014 to 62% in 2016. Real Estate and Infrastructure were the two sectors most commonly targeted by Sovereign Wealth Funds. ¾ of the Sovereign Wealth Funds investing in infrastructure have a separate allocation for this asset class, often targeted at 5% of their portfolio, and currently around 3% of their total asset under management.

However, in parallel, international foreign investments in key national assets still inspire fear. Xu (2009) spoke of "the spectre of a rich Chinese state buying strategic resources, hollowing out companies, gobbling up financial institutions and threatening the sovereignty of the countries in whose resources and companies it invests." This statement is still relevant today: the lease of the port of Darwin, Australia to a Chinese company in 2015 started a geopolitical crisis with the USA. The press and military experts underlined the potential threat to the US military base in Darwin, and to the US military network in Asia-Pacific.

Sovereign Fund (Government-owned fund investing on private markets for financial returns) are both wanted and unwanted as infrastructure investors. It is therefore more relevant than ever to understand what type of investors they are. It also builds on a long tradition. The issue of trust and the legitimacy of Sovereign Funds as international investors dominated academic discussions between 2007 and 2012. Many analyses explained these fears as a deeper change in the world geopolitical balance, and a clash of models of capitalism, between the free market and state capitalism (Monk, 2008; Santiso, 2008; Drezner, 2008).

However, in parallel, a stream of literature spearheaded by the OECD's long-term investing group emphasized the benefits that this growing group of investors could bring to national economies. Sovereign Funds' long-term horizons and development goals made them ideal investors in SMEs, infrastructure, innovation, and other sectors where private investors were scarce, but the needs acute.

We argue that the growth in publications and intense debate about Sovereign Funds' identity has given them the opportunity to become their own category. A large literature both in economics, economic geography, political economy, and management emphasized the need for these funds to increase their transparency to showcase their non-political motives, good governance and financial performance. Alongside this literature, international organizations such as the World Economic Forum and the OECD started working groups on Sovereign Wealth Funds. They brought together these funds, published surveys and principles of good governance, all to lead these funds to be more transparent. Today, most sovereign funds have professional websites, publish annual reports, and sometimes even responsible investment reports, their own investment principles, quarterly reports, reports to stakeholders, and other documents, all building their public persona.

Interestingly, the search for transparency can have two opposite interpretations. First, it can be a way to integrate these funds into the existing field of private investment funds, by forcing them to adopt the same codes, values and processes and steer them away from their political and social goals. On the other hand, it can be an opportunity for these funds to present themselves as Sovereign Investors per se, a new category of investors, with different missions an investment principles than private investors <u>and</u> from government. The third alternative is that we see

subgroups appearing in a category that still seems artificial to some of these funds. Indeed, pension funds and sovereign funds are often not bound by the same laws and regulations, and have different financial liabilities, perhaps justifying clearly differentiated identities.

Most of the academic research on Sovereign Funds so far had to be built on case studies and surveys, which limited our understanding of how much the findings were representative of the whole population. Today's almost unanimous production of websites and annual reports by Sovereign Funds gives an opportunity to do a large-scale comprehensive study of these actors. It also enables us to use primary data, produced by these actors, instead of descriptive data gathered by observers.

We aim to distinguish groups of sovereign investors based on the topics and vocabulary they use to describe themselves. The resulting taxonomy will call into question the numerous criteria used to classify these funds: which one explains best the differences in declared identity of these funds?

This paper analyzes the literature on sovereign funds and make the case for an analysis of annual reports over the past ten years using computer-assisted text analysis to answer these questions. We also propose a scope and methods of analysis but leave the actual analysis and results for a future paper.

HOW DO WE ANALYZE AND CATEGORIZE SOVEREIGN FUNDS TODAY?

"Sovereign wealth funds" (SWFs) were first defined by Rozanov in 2005 as "neither traditional public pension funds nor reserve assets supporting currencies, but a different type of entity altogether." In 2008, Monk noted that 3 years later, there was no consensus on how to define SWFs and how they differed from public investment funds. A debate also included if they could be considered by foreign governments and international laws in the same way as private investment funds (Backer, 2008). What seemed constant was the analysis that Sovereign Funds should be recognized as a new category of investors, at the intersection of "high finance and high politics" (Drezner, 2008; Monk, 2008). SWFs were analyzed as a manifestation of the financialization of government, and a geopolitical reversal in which many developing countries that used to be debtors were becoming creditors and now investing in OECD countries.

A field of research developed around the puzzle of SWFs being in between two institutional fields: the state and international finance. This led to uncertainty about the category of organizations they were. Definition proposed emphasized this position at the intersection of two fields. Monk (2008) proposed the following definition: "*government-owned investment funds operating in private financial markets.*"

Drezner's definition (2008) also shows that SWFs are inside the field of government but operate in spaces traditionally reserved to private investors and according to the same goals as private investors: "*government investment vehicles that acquire international financial assets to earn a higher-than-risk-free rate of return.*"

The question many economists, lawyers and politicians[3] asked was: can a government-owned fund invest like a private investment organizations, or are they trojan horses, penetrating western markets to better destabilize some countries for political reasons?

---

[3] See Monk's interviews at the White House in his 2008 paper

THE CALL FOR SWFS TO BEHAVE LIKE PRIVATE INVESTORS: A STORY OF CONFORMITY FOR LEGITIMACY

Taking an economic perspective, Kotter & Lel (2011) cite literature showing that government-ownership comes with potential political interference, agency conflicts, poor financial performance, inefficiency and weak managerial incentives. They propose that government's non-financial goals, no matter if they are political or social, and aggressive toward a country, or wishing to develop one's country, are incompatible with the interest of the companies in which SWFs invest. They also posit that transparent SWFs are more likely to be financially-oriented, whereas opaque SWFs are more likely to be political vehicles. The authors show that SWFs' preference for distressed companies in developed countries, and effect on these firms' financial performance in the long-run are similar to other passive investors, reinforcing the idea that SWFs act as private investors. However, the authors also show that transparent SWFs have a more positive market reaction when they buy stocks, which leads the authors to state that "they provide evidence in support of policies recently initiated by SWFs aimed at improved transparency", namely the Santiago Principles.

From a legal perspective, Backer (2010) analyzed the Norwegian Government Pension Fund as an example of the socially responsible SWF. He concludes that even if SWFs say they function and act like private investors, their behavior is equivalent to indirect regulation. By prohibiting investments in countries or companies that are not socially responsible, they are a "vehicle for extraterritorial application of municipal law". Backer's proposed solution is for SWFs to be functionally distinct from the state, and to strictly behave like an idealized private investor. He notes that the Santiago Principles theoretically achieve both, even though his case study of Norway shows that in practice, the Nordish SWF does not behave like an idealized private investor, and that their macro-economic and ethical missions drive behaviors that are State-like.

The Santiago Principle are a key representation of the will to softly regulate SWFs. They were drafted by 14 funds forming the International Forum of Sovereign Wealth Funds in 2009, as part of a joint effort between the international monetary fund and the "international working group of sovereign wealth funds" to bring more transparency into the governance and investment decisions of SWFs. The signatories of the Santiago Principles, now 30 SWFs, are members of the IFSWF and publish yearly reports about how they are working toward implementing the Santiago principles.

Effectively, the Santiago Principles are a declaration by these funds that they will abide by the rules of international financial markets. One of these rules is to invest for financial returns exclusively. Another important one is to be transparent about the fund's mission, investments and financial performance.

This effort and the effective increase in transparency of SWFs as measured by the publication of annual reports in English can be interpreted as the legitimation of SWFs in the eye of Western democracies. It aimed at aligning their governance and investment practices with neo-liberal capitalist expectations and norms for private investors (Monk, 2009). It goes hand in hand with the trend of financialization of the State (Monk, 2009). It also portrays the global trend of rationality and tendency to make accountability equal auditability described by Meyer as part of his theory of Modernity (Bromley and Meyer, 2014; Meyer's book).

However, in parallel to the creation of the IFSWF and the Santiago Principles, the OECD created a working group for long-term investing, bringing together pension funds and SWFs around new principles for investment.

THE OPPORTUNITY FOR SWFS TO BECOME THEIR OWN CATEGORY AND BUILD THEIR FIELD ON NEW INVESTMENT VALUES

The call for SWFs to behave exactly like private investors happened at the same time as the Global Financial Crisis of 2008, a major jolt in the international economy. The GFC both dried up funding for many sectors known to support the economy (Infrastructure, SME, PE), and reinforced voices condemning the short-term horizons of private investors and their role as detractors of financial markets. The GFC can therefore be considered as a perturbator of the field of global finance and its message that pursuing pure financial returns is the best way to support the economy. Such events have been described as essential to the emergence of a new institutional field, in the sense that a new set of actors can now propose a new set of rules and values, and reorganize relations between actors (Fligstein & McAdam, 2012, Meyer et al., 2005, Meyer, 1982). It reinforced a competing institutional logic emphasizing the need for socially-responsible companies and investing, and potentially weakened the power of the field of finance to influence the emerging actors that SWFs were (Meyer, Pope & Isaacson, 2014).

In this context, governments started seeing some institutional investors as a potential solution to the flaws of the financial system. The OECD and the G20 proposed in 2013 the High-Level Principles of Long-term Investment Financing by Institutional Investors, almost as a parallel of the Santiago Principles, but now advocating for a new identity and role for a group of institutional investors.

First, this initiative broadened the boundaries of the group of Sovereign Investors. While the Santiago Principles were targeted at Sovereign Wealth Funds, the long-term investment principles (LTIP) were targeted at a larger group of institutional investors including Public Pension Funds (PPFs), Sovereign Wealth Funds (SWFs) and Insurance companies. Before this initiative, the literature on SWFs had drawn a line between the two: SWFs did not have fixed liabilities, whereas PPFs had this mandatory future stream of payments (Monk, 2011). By 2016, PwC's working group on SWFs published a report about "Sovereign Investors" including public pension funds and sovereign development funds in this definition.

The high-level principles for LTI were also the sum of a work bringing together many actors of what could be considered an emerging institutional field: 4 formal talks and written contributions from the OECD, G20, FSB, APEC members, members of the Task Force, several OECD bodies (such as the Committee on Financial Markets (CMF) and the Insurance and Private Pensions Committee (IPPC)), International Organizations (IMF, World Bank, FSB, various SSBs), G20 Study Group on Finance for Investment and the European Commission. The aim was to change countries' regulations to favor long-term investing practices, which effectively was a move supporting the institutionalization of a new group of investors: "The high-level principles are designed to assist OECD, G20 and any other interested countries to facilitate and promote long-term investment by institutional investors, […] foster consistency in approaches for long-term investment across different policies and jurisdictions." (OECD, 2013)

This document and the series of reports on sovereign funds, pension funds and long-term investing produced by the OECD/G20 taskforce and the World Bank (Gelb et al., 2014) make investing for

development legitimate. Sovereign investors are not dangerous anymore, but "bring about the prospect of a larger and more diversified source of long-term financing for physical and intangible investment needs across all sectors in the economy and specifically in key drivers of growth, competitiveness and employment such as infrastructure, company equipment, education and skills, research & development, and new technology." The principles also propose new key words for these investors. They are characterized as "'patient', 'productive' and 'engaged' profitable capital" (OECD, 2013). "Patient" refers to their structural characteristics as strategic advantages for them to reap higher financial returns: illiquidity premia, lower turnover, less pro-cyclical investment, higher net investment rate of returns and greater financial stability. "Productive" refers to asset classes where these investors can leverage their characteristics: infrastructure, green growth, SMEs. "Engaged" refers to active voting policies leading to better corporate governance, which is the contrary to what had been observed in economic studies such as Kotter et Lel's (2008).

The Principles of LTI in fact redefine investing for "political and social goals" into a legitimate form of investment in line with the global trend requiring transparency and rational, financial-return seeking behaviors, but crafting a space where development goals are no longer antithetic with it. Structural characteristics of institutional investors are used to create the idealized "long-term investor", and align it with government's development goals.

RESEARCH QUESTIONS

The overarching goal of this research is to contribute to a more detailed understanding of what makes Sovereign Funds a new category of investors, and question the boundaries and internal consistency of the category. Therefore, we ask the following questions:

1. What self-chosen vocabulary and themes enable to differentiate between Sovereign Investors?
2. How do the groups based on these themes map unto commonly used criteria to categorize sovereign investors?

To answer these questions, we gathered annual reports since 2004 (when available) for organizations that have been classified as Sovereign Funds. We detail in the next part how we selected these organizations.

SCOPE OF THE STUDY: CHOICE OF ORGANIZATIONS TO INCLUDE IN THE STUDY

Sovereign Funds, also called sovereign investors, constitute the population of interest. This category has been increasingly used since 2007 by funds themselves and by observers, international organizations and consulting groups. The longest list of SFs we found comprises 129 entities and was proposed in 2016 by PwC in their report *Sovereign Investors 2020*. We compared and completed this repertory with the list of sovereign funds, and government and sovereign institutions from the Sovereign Wealth Center.

We also included members of international organizations dedicated to government-owned investors and long-term institutional investors. Monk, Sharma, and Sinclair (2017) lists 7 roundtables and research clubs for long-term investors. Among those, we used the International Forum of Sovereign Wealth Funds, the Long-Term Investment Club, the Institutional Investors Roundtable, and the Forum Mondial des Caisses de Depots to identify organizations that would be

considered SFs. They also include the OECD long-term investment platform and the WEF Long-term Investment Council. These organizations are mainly research-oriented, and do not really provide a list of funds, but they work hand in hand with the IIR, the LTIC and the IFSWF, and we could use their publications as part of potential influencers of the topics appearing in SF's annual reports. Monk et al. also cite the Pacific Pension & Investment Institute. This organization lists asset owners, sovereign funds but also cities', counties' and private companies' pension funds, which are outside our population. However, they also have a list of asset managers, who could be our control group when it comes to comparing sovereign funds and private investment funds. In the same vein, the asset owners specific to the PPI and not listed by other sources could be a control group representing other types of pension funds.

Indeed, professional associations and research centers are a carrier of values and key words that are then used by an increasing number of organizations around the globe (Meyer, Pope, Isaacson, 2014; Greenwood et al., 2002). They are motivated by mimicry instead of specific practical needs. A common vocabulary characterizes a new category of organizations. Here, one could expect all organizations in the PPI to slowly adopt the same vocabulary, or for some groups to form around their own vocabulary, distinct from other sub-groups inside the PPI. Showing different sub-groups based on text analysis would strengthen the hypothesis that SFs are their own category of investors, with the potential to impose a new set of values and practices in the investment world.

Another emerging subgroup not represented in Monk et al.'s list is that of Sovereign Development Funds, also called Strategic Funds (Halland et al., 2016). We therefore included 12 funds listed in an article on Sovereign Development Funds by Clark & Monk (2015) that did not completely overlap with the list coming from the sources listed above. Appendix 1 gives more details about each source.

The boundaries of the category "sovereign investors" are not well defined and **table 1** shows that the degree of overlap between the different groups is often quite low, suggesting that bringing the SFs from these groups might show differences between groups, or an interesting uniformity in language and identity, potentially thanks to the organizations being members of several groups. The numbers in the top right corner may seem dramatic, but reflect in great part the fact that these groups are small compared to the full list of SIs according to SWC and PwC. However, it is surprising that even though the SWC and PwC list a large number of organizations, they miss such a large number of the organizations that are part of the LTIC and FMCD, as is shown in the bottom-left corner. It could indicate a geographical or cultural discrepancy, since the IFSWF, PwC and the SWC are primarily USA and Great Britain based, while the LTIC is a European initiative, and the FMCD is widely francophone. Finally, even if the overlap between SWC and PwC is better than for others, it is still less than ¾, which is low for two initiatives that aim at representing the full scope of SIs.

Table 1: Overlap table

Indicates the % of sovereign funds listed in "organization in row" that are also in listed in "organization in column". The diagonal indicates the number of funds listed in the organization, e.g. Sovereign Wealth Center lists 97 sovereign funds. 72% of these sovereign funds are also listed by PwC, while only 9% of them are part of the IIR.

	SWC	PwC	C&M	IFSWF	LTIC	FMCD	IIR
Sovereign Wealth Center (SWC)	97	72%	8%	27%	27%	4%	9%

Sovereign Investors 2020, by PwC 2016 (PwC)	59%	119	8%	24%	5%	3%	6%
Clark & Monk 2015 (C&M)	67%	67%	12	33%	8%	0%	8%
International Forum of Sovereign Wealth Funds (IFSWF)	84%	90%	10%	31	16%	3%	23%
Long-term investors club (LTIC)	32%	32%	5%	26%	19	21%	21%
Forum Mondial des Caisses de Depots (FMCD)	29%	29%	0%	7%	29%	14	21%
Institutional Investors Roundtable (IIR)	90%	70%	10%	70%	40%	30%	10

From the eight sources we mentioned above, we have a list of 165 funds that could classify as sovereign investors according to at least one source. No organization belongs to more than 5 funds. See **table 2** for how many funds are cited by 1, 2, 3, 4 and 5 sources. Since the number goes down so quickly, we decided to start with funds that belong to 3 groups or more to keep diversity but make the data collection more manageable. This gives us 35 organizations for which we collect the annual reports over the past 10 years. We use the websites of the funds when they do not publish annual reports, and consider it for only one year (indicated on the website).

Table 2: Number of Funds cited in 1, 2, 3, 4 and 5 groups.

1 group	2 groups	3 groups	4 groups	5 groups
80	50	29	3	3

The uncertainty around who is a sovereign investor, and the potential geography-based differences make the question of uniformity of language among these organizations even more interesting. We detail below our hypotheses as well as the methods we think would most effectively question current categories.

HYPOTHESES

Based on the literature on Sovereign Investors, we proposed the following two main hypotheses.

Hypothesis 1: presence of development goals in annual reports

Meyer et al. (1997) showed that field logics have often spread and institutionalized through international organizations, professional associations and academia. We have delineated a neo-institutional story showing how two institutional logics have played a role in the emergence of the category of sovereign investors. Our first hypothesis seeks to highlight how sovereign investors position themselves regarding investing for other reasons than pure financial returns. If there is emergence of a new field logic, annual reports of sovereign investors would give a place to development goals, and potentially reuse the vocabulary and intent proposed by the OECD, which seek to legitimate investing for development. Alternatively, sovereign investors might purposely downplay their development mission in their annual report, making them look and sound as similar as those of private investment funds as possible.

Hypothesis 2: AUM, Region, and Professional Groups are the most informative criteria

First, pension funds were added to the category of sovereign investors quite late in the emergence of the category. Many of them are professional investors, with large assets under management, and a growing exposure to real assets, which can justify their belonging to this category. We propose to check if pension funds, sovereign wealth funds and sovereign development funds differ in terms of topics mentioned in their annual reports. We assume that pension funds might emphasize, safety, reliability, passive investment more than sovereign funds. We also expect sovereign development funds to emphasize local initiatives, development, political aspects and social initiatives more than sovereign wealth funds and pension funds.

We also propose to use the following criteria as representative of sub-categories and groups of Sovereign Wealth Funds:

- Pension funds, sovereign wealth funds, sovereign development funds
- Assets under management (AUM)
- Region: Asia, Pacific, Africa, Americas, Europe, Middle-East
- Age
- Belonging to a professional association

Monk and Dixon (2016) recently called for more "mission clarity" as key to the success of individual sovereign funds. They emphasized that "Sovereign funds cannot be all things to all people", and that specific missions called for different investment principles and portfolios. The authors use the mandate as a good proxy for the identity: stabilization fund, long-term savings, and catalytic development.

Based on the literature on sub-categories, we propose the following hypothesis regarding how annual reports might be influenced by these criteria.

Hypothesis 2a: The size of AUM and the region of origin are adequate criteria to group SIs and inform the taxonomy based on topics.

Hypothesis 2b: Pension funds, sovereign wealth funds, sovereign development funds do not match well with the identities based on topics.

Hypothesis 2c: Membership in an association is associated with being part of the same group based on topics in annual reports.

METHODOLOGY

We review first how annual reports have been used to analyze organizational and corporate identities in previous literature, before showing the opportunity that computer-assisted methods of text analysis and topic modeling have offered in other fields, notably in political science.

USING ANNUAL REPORTS TO ANALYZE ORGANIZATIONAL IDENTITY

Organizational identity can be defined scientifically based on the organization's characteristics to an external observer, or internally, by the characteristics that the organization gives to itself. Both definitions assumes that there are essential features forming the central character of an organization, that are distinctive from other organizations and show a degree of sameness over time (Albert & Whetten, 1985). A central question is the difference between how an organization presents itself, and how it actually functions and behaves. Albert & Whetten (1985) underlines organizations'

definition of themselves ss a political, strategic act, differing on the audience and purpose of the declaration. However, the way the organization invokes and creates classification schemes and locates itself in comparison gives interesting information, because it positions the organization in a field of other actors.

The chosen categories also help or hinder an organization's relationships with other actors, and most notably its clients and partners. Durand and Khaire (2016) distinguish between category emergence and category creation, underlining both the fact that organization need to react to some categories coming into their field, but can also push for the creation of a new category as a strategic advantage. In this case, the legitimation of SFs as a category can calm fears of undue state influence thus making investment partnerships easier.

Recognizing that communication around their identity is strategic, studying this controlled communication will not necessarily be representative of actual investment strategies and internal logics. However, it gives access to the process of sense-making that has helped to understand product-markets notably (Weick, 2005; Rosa et al. 1999). The vocabulary and categories influence relationships between these funds and governments, their peers, and their competitors.

The choice of using annual reports is also strategic. These documents are easily accessed by the public, and their target audience is very large: local actors, potential partners, foreign governments, investors, potential investee, etc. They have also been used by private companies for decades and are now highly institutionalized. Their plan and the financial information one expects in them are widely similar for all organizations, and even more similar in one industry. However, there is also space to affirm a company's identity, notably in the design, color-code, images, slogan, and a few key words representing priorities and values. They are therefore an exercise in legitimacy and positioning among competitors. Mission statements inside these reports are particularly important (Leuthesser & Kohli, 1997). In this study, we don't discuss how close to reality these statements and reports are, or if they have a goal-setting value, but take them as an expression of what they think will bring them legitimacy and/or competitive agency, and how close their language effectively is to other organizations' (Argenti & Druckenmiller, 2004; Esrock & Leichty, 1998).

To get at an organization's identity among existing categories, Albert & Whetten envisioned a method in which listing all the characteristics of several institutionalized types of organizations, and positioning organizations on this spectrum can help define categories and organizations' identities. This method faces the problem of how one puts together a non-biased, complete list of these characteristics, and match it with the vocabulary used by organizations. This is where techniques of computer-assisted text analysis bring both faster and more reliable analyses of text, and the possibility to find new topics and groups of organizations.

COMPUTER-ASSISTED TEXT ANALYSIS AND TOPIC MODELING

Computer-assisted text analysis are increasingly used for research in social sciences. Indeed, these methods greatly reduce the cost and time needed to analyze large collections of text. (Grimmer & Stewart, 2013) They can be separated in two broad categories: supervised and unsupervised, based on the degree of intervention of the analyst. The number of methods is already extremely large and expanding, and it is beyond the scope of this paper to review them (see Grimmer's project to provide a software, consilience.com, to allow researchers to compare the results from a large number of these methods (Grimmer, 2010)). We present a few limitations, response to them, and the methods we propose to use for our study.

We are interested both in categorizing documents based on hypotheses, and to try and see if topics would emerge based on learning from key texts and on metadata instead of coding directly the documents. The first method replicates the manual coding task, but with a machine. It usually uses a 'dictionary', meaning a set of words that correspond to a category. It can be more or less guided. First, we could just represent the frequency of words related to development versus financial returns, which is very bounded. Second, we could leave more possibilities by training an algorithm on texts from the OECD and the Santiago Principles as representative of two categories, and using this algorithm to classify annual reports. Both methods are supervised: they involve explicitly setting the categories based on models and/or a dictionary, and assigning documents to a predetermined categorization scheme (Grimmer & Stewart, 2013). We plan to validate these methods by reading documents coded in each category and checking for consistency.

Another way to categorize the annual reports is to use unsupervised methods, such as topic modeling. In the words of Grimmer & Stewart (2013) "Unsupervised learning methods are a class of methods that learn underlying features of text without explicitly imposing categories of interest. Rather than requiring users to condition on known categories beforehand—supervising the methods—unsupervised learning methods use modeling assumptions and properties of the texts to estimate a set of categories and simultaneously assign documents (or parts of documents) to those categories." Using this type of methods would help us discover new clusters of documents and new topics that we did not posit beforehand. It is even more interesting because the category is emerging and new models based on a set of principles and words might be gaining ground even though they are not explicitly pushed by international organizations. Indeed, our hypotheses are based on a top-down view of institutionalization. However, a bottom-up view also exists, and unsupervised methods could help uncover topics unidentified in our hypotheses. Using topic modeling could help show if regional proximity, categories such as pension funds, reserve funds, or capital maximization funds do correspond with proximity in topics. This is now feasible using topics modeling algorithm that incorporate metadata such as the Structural Topic Modeling (stm) method and associated package in R. This method builds on the Latent Dirichlet Method and allows researchers to discover topics and estimate their relationship to document metadata (Roberts, Stewart, Tingley, 2015; structuraltopicmodel.com; Lucas et al., 2013).

We detail in appendix 2 the existing criteria used in the literature to create subcategories of funds, and which one we will use as metadata.

Grimmer & Stewart (2013) emphasize that "all quantitative models of language are wrong but some are useful". Indeed, the methods we presented above do not understand text the way humans do. They compute occurrences of words, or bigram and trigrams as put in a document term matrix by the researcher. They do not replace humans, and there is no one globally best method. The researcher is therefore still very present in selecting the method, how to use it, and how to interpret it. Therefore, validation of the results given by the technique is essential. We plan to validate the results by reading a selection of the annual reports to validate our interpretations. We do think that these methods will guide a different lecture of these annual reports and provide a methods of categorizing sovereign funds that is both quantified and grounded in the words of the organizations studied.

EXPECTED CONTRIBUTIONS:

Sovereign funds are increasingly presented as the solution to the lack of financing of infrastructure projects. However, they have faced institutional challenges to becoming a new breed of investors, both in terms of regulations and cultural frameworks. We argue that sovereign funds are an emerging category of investors, still searching for an identity that could support a double bottom line approach.

The setting described above offers an opportunity to study the emergence of a category of organizations, looking at how key terms and topics move from some actors to others, as well as how close or similar these SFs are to institutionalized actors (private institutional investors and government agencies notably). This paper laid out that the literature on sovereign funds has been prescriptive so far. We now data to find categories from the ground up.

We call for the use of computer-assisted text analysis to broaden the scope of the research on sovereign funds, and give more weight to the words used by these organizations. Such an analysis, on which we are currently working, will give another basis to the categories that have been proposed to date. It will also anchor these categories in the missions and declared identities of these organizations.

Practically, a better understanding of the landscape of sovereign funds can help sovereign funds know who are the closest culturally, and support their choice of partners. It can also help infrastructure project managers choose partners. It also shows which countries and funds claim values in line with the needs of infrastructure. This taxonomy can also help choose case studies for future research on the alignment between declared identities and actual investment behaviors.

Future research could use longitudinal data to study the emergence of new sub-groups of sovereign investors through the evolution of their vocabulary. It could also test the influence of observers, professional organizations and professional media on how organizations in these groups present themselves to the public and potential partners and competitors.

LITTERATURE CITED AND PROPOSED FOR FUTURE DEVELOPMENT OF THE PAPER

Albert, S., & Whetten, D. (1985). Organizational Identity. Research in Organizational Behavior, 7, 263–295.

Alsweilem, K. A., Cummine, A., Rietveld, M., & Tweedie, K. (2015). A comparative study of sovereign investor models: Sovereign fund profiles, 1–142. Retrieved from http://projects.iq.harvard.edu/sovereignwealth/comparative-study-sovereign-investor-models-sovereign-fund-profiles

Argenti, P. A., & Druckenmiller, B. (2004). Reputation and the Corporate Brand. Corporate Reputation Review, 6(4), 368-374

Ashforth, B., & Mael, F. (1989). Social Identity Theory and the Organization. The Academy of Management Review, 14(1), 20-39. Retrieved from http://www.jstor.org/stable/258189

Bachher, J. S., & Monk, A. H. B. (2012). Attracting Talent to the Frontiers of Finance.

Backer, L. C. (2009). Sovereign Wealth Funds As Regulatory Chameleons: the Norwegian Sovereign Wealth Funds and Public Global Governance Through Private Global Investment.

Georgetown Journal of International Law, *41*(2), 425. Retrieved from https://ssrn.com/abstract=1398835

Balding, C. (2009). Who's Afraid of Sovereign Wealth Funds? *Revue D'economie Financiere*, 201–210.

Bromley, P., Meyer, J. (2014) "They Are All Organizations" The Cultural Roots of Blurring Between the Nonprofit, Business, and Government Sectors - Administration & Society

Brown, A. (1997). Narcissism, Identity, and Legitimacy. The Academy of Management Review, 22(3), 643-686. Retrieved from http://www.jstor.org/stable/259409

Chreim, S. (2005). The Continuity--Change Duality in Narrative Texts of Organizational Identity. Journal Of Management Studies, 42(3), 567-593. doi:10.1111/j.1467-6486.2005.00509.

DiMaggio, P., Powell, W. W. (1983) "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality" in Organizational Fields, American Sociological Review 48: 1983: 147-60

Dixon, A., Monk, A. (2012) "Reconciling Transparency and Long-Term Investing within Sovereign Funds." Journal of Sustainable Finance & Investment: 37–41.

Dixon, A., Monk, A. (2014) "Financializing Development: Toward a Sympathetic Critique of Sovereign Development Funds."

Dixon, A. D., & Monk, A. H. B. (2016). A Simple Typology of Sovereign Development Funds. In The Frontiers of Sovereign Investment. Columbia Press.

Duriau, V. J., Reger, R. K., Pfarrer, M. D. (2016) A Content Analysis of the Content Analysis Literature in Organization Studies: Research Themes, Data Sources, and Methodological Refinements, Organizational Research Methods, Vol 10, Issue 1, pp. 5 – 34

Esrock, S. L., Leichty, G. B. (1998) Social responsibility and corporate web pages: Self-presentation or agenda-setting? Public Relations Review, Volume 24, Issue 3, Pages 305–319.

Fligstein, N., & McAdam, D. (2012). "3. Macroconsiderations" in *A theory of Fields*

Gelb, A., Tordo, S., Halland, H., Arfaa, N., Smith, G., (2014) "Sovereign Wealth Funds and Long-Term Development Finance, Risks and Opportunities", Policy Research Working Paper 6776, The World Bank

Greenwood, R., Suddaby, R., & Hinings, C. (2002). Theorizing Change: The Role of Professional Associations in the Transformation of Institutionalized Fields. *The Academy of Management Journal, 45*(1), 58-80. Retrieved from http://www.jstor.org/stable/3069285

Grimmer, J. (2013) Representational Style in Congress, What Legislators Say and Why It Matters, Stanford University Press

Grimmer, J., Stewart, B. (2013) "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts", Political Analysis, 21:267–297 doi:10.1093/pan/mps028

Grimmer, J. (2010) "Representational Style: The Central Role of Communication in Representation", Dissertation, Harvard University

Halland, H., Noël, M., Tordo, S., Kloper-Owens, J. (2016) "Strategic Investment Funds, Opportunities and Challenges", Policy Research Working Paper 7851

Haniffa, R., & Hudaib, M. (2007). Exploring the Ethical Identity of Islamic Banks via Communication in Annual Reports. Journal of Business Ethics, 76(1), 97-116. Retrieved from http://www.jstor.org/stable/25075497

Keller, A. (2009). Sovereign Wealth Funds: Trustworthy Investors or Vehicles of Strategic Ambition ? An Assessment of the Benefits, Risks and Possible Regulation of Sovereign Wealth Funds, (May), 333–372.

Kotter, J., & Lel, U. (2011). Friends or foes? Target selection decisions of sovereign wealth funds and their consequences. Journal of Financial Economics, 101(2), 360–381. http://doi.org/10.1016/j.jfineco.2011.03.007

Leuthesser, L., & Kohli, C. (1997). Corporate identity: The role of mission statements. Business Horizons, 40(3), 59.

Lounsbury, M. (2002). Institutional Transformation and Status Mobility: The Professionalization of the Field of Finance. *The Academy of Management Journal, 45*(1), 255-266. Retrieved from http://www.jstor.org/stable/3069295

Lucas, C., Nielsen, R. A., Roberts, M. E., Stewart, B. M., Storer, A., & Tingley, D. (2012). Online Appendix: Additional Text Preparation Details and A Review for Text Analysis Techniques for Comparativists. Political Analysis, 23(2), 1–23.Meyer, A.D., Gaba, V., Colwell, K. A. (2005) "Organizing Far from Equilibrium: Nonlinear Change in Organizational Fields" Organization Science 16(5):456-473.http://dx.doi.org/10.1287/orsc.1050.0135

Meyer, A.D., (1982) "Adapting to Environmental Jolts", Administrative Science Quaterly 25, 515:539

Meyer, John W. and Patricia Bromley (2013) "The Worldwide Expansion of 'Organization'", Sociological Theory 31: 4: 343-365.

Meyer, John W., David Frank, Ann Hironaka, Evan Schofer, and Nancy Tuma "The Structuring of a World Environmental Regime, 1870-1990." International Organization, 51, 4, Autumn 1997: 623-51

Meyer, J. W., Pope, S., and Isaacson, A. (2014) "Legitimating the Transnational Corporation in a Stateless World Society" In K. Tsutsui and A Lim, eds., Cambridge

Monk, A. (2009). Recasting the Sovereign Wealth Fund Debate: Trust, Legitimacy, and Governance. New Political Economy (Vol. 14). http://doi.org/10.1080/13563460903287280

Monk, A. H. B. (2011). Sovereignty in the era of global capitalism: the rise of sovereign wealth funds and the power of finance. *Environment and Planning A*, *43*(8), 1813–1832. http://doi.org/10.1068/a43326

Monk, A., Sharma, R., Sinclair, D. (2017) Reframing Finance, New Models of Long-Term Investment Management, Stanford Press.

Musacchio A., Lazzarini, S. (2014) "Reinventing State Capitalism, Leviathan in Business, Brazil and Beyond", Harvard University Press.

Navis, C., & Glynn, M. A. (2010). How New Market Categories Emerge: Temporal Dynamics of Legitimacy, Identity, and Entrepreneurship in Satellite Radio, 1990-2005. Administrative Science Quarterly, 55(3), 439-471

PwC (2016) "Sovereign Investors 2020"

Rozanov, A. (2005), "Who Holds the Wealth of Nations", Central Banking Quarterly Journal, Vol. 15, no. 4

Suddaby, R., Greenwood, R. (2005). "Rhetorical Strategies of Legitimacy" Administrative Science Quarterly, 50 (1), 35-67. Retrieved from http://www.jstor.org/stable/30037175

"Managing Managerial Identities: Organizational Fragmentation, Discourse and Identity Struggle", Stefan Sveningsson, Mats Alvesson, Human Relations, Vol 56, Issue 10, pp. 1163 − 1193, First published date: April-22-2016, DOI: 10.1177/00187267035610001

Ullmann, A. (1985). Data in Search of a Theory: A Critical Examination of the Relationships among Social Performance, Social Disclosure, and Economic Performance of U. S. Firms. The Academy of Management Review, 10(3), 540-557. Retrieved from http://www.jstor.org/stable/258135

Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (2005). Organizing and the Process of Sensemaking. *Organization Science*, *16*(4), 409–421. http://doi.org/10.1287/orsc.1050.0133

Xu Y. (2009). "Who's afraid of sovereign wealth funds? *Australian Journal of International Affairs*, *63*, 1–21. http://doi.org/10.1080/10357710802666141

Robbins, S. S., Styliano, A. (2003) "Global corporate web sites: an empirical investigation of content and design" Information & Management, Volume 40, Issue 3, Pages 205–212

Roberts, M. E., Stewart, B. M., & Tingley, D. (2015). stm: R Package for Structural Topic Models. *Journal of Statistical Software*, *VV*(2014).

APPENDIX 1

NAME	FOUNDING/ PUBLICATION DATE	# OF FUNDS LISTED	SHORT DESCRIPTION OF THE FUND AND RELEVANCE TO THE TOPIC (BASED ON THE WEBSITES OF EACH ORGANIZATION, CONSULTED IN APRIL 2017)
SOVEREIGN WEALTH CENTER (SWC)	2012	97	Database of funds created by the periodical institutional investors to provide market-leading research on SWFs to asset managers, investment banks and consultants. Their goal is to "help the wider investment community keep abreast of the latest investment trends in this important group of institutions". SWC is also part of the sovereign investor institute, a global forum of sovereign and government funds, created in 2010 and with 6 roundtables annually. SII does not have a website or a publicly available list of members
SOVEREIGN INVESTORS 2020, REPORT BY PWC 2016	2016	117	PwC published their first report on SF in October 2013, focusing on the role of SFs in the board room. They now have a team dedicated to SFs called the "SIF leadership", working with "the global team" to publish reports on SFs. In 2016, they published a comprehensive review and categorization of SFs, notably distinguishing between pension funds and sovereign wealth funds. They also gave prospective numbers about the growth of these funds, and a perspective on how global megatrends affected and were affected by SFs.
CLARK & MONK, 2015	2015	12	This article is a comparative case study of SFs with a primary mission of national development. SDFs have not really made it yet into the vocabulary of consulting firms and observers, but "strategic investments" aimed at developing a specific region is increasingly used. Few funds call themselves sovereign development funds, and it is unclear how widespread the declaration of having a development goal is. The present article aims at clarifying this point, which is why we tried to include funds identified as SDFs in the literature.
INTERNATIONAL FORUM OF SOVEREIGN WEALTH FUNDS (IFSWF)	2009	30	The IFSWF came out of a joint effort between the international monetary fund and the "international working group of sovereign wealth funds to bring more transparency into the governance and investment decisions of SWFs. They drafted the 24 Santiago Principles and brought together major SWFs to sign them and join the IFSWF. The members publish yearly reports about how they are working toward implementing the Santiago principles.
LONG-TERM INVESTORS CLUB (LTIC)	2009	19	The LTIC was founded in 2009 by a French and an Italian sovereign fund, the German development bank and the European development bank, and grew to have 19 members in 2016. LTIC's mission is to bring together major worldwide institutions and emphasize their common identity as long-term investors, as well as to encourage cooperation and common long-term investments promoting growth. Two equity investment funds bringing together members have come out of the LTIC. The LTIC also produces publication and works toward a common language and identity around the theme of development and long-term investment.
FORUM MONDIAL DES CAISSES DE DEPOTS(FMCD)	2011	14	FMCD holds large conferences every 2 years (2011, 2013, 2015 and 2017). The third one was in Tunisia and focused on the themes of support to small and medium enterprises, infrastructure financing, and housing and urban development. Until spring 2017, the Cassa di Depositi e Prestiti (Italy), the CDC (Tunisia), the Caisse des dépôts (France), and the CDG (Morocco) formed the general secretariat. This organization shows a clear focus on development and on a type of government-owned investment funds called Caisse des dépôts. These organizations are very similar to sovereign funds and sometimes call themselves sovereign funds but are less often identified and categorized as such by (almost exclusively American and British) research centers on sovereign funds.

INSTITUTIONAL INVESTORS ROUNDTABLE (IIR)	2011	10	Non-profit organization founded by 10 sovereign funds with the goal of improving collaboration among sovereign funds, foster trust and envision ways to collaborate and co-invest directly, notably in private equity. The IIR holds two annual meetings open to a larger community of sovereign funds. We only list here the founding members, involved in managing the platform and organizing the conferences.

APPENDIX 2: SUB-CATEGORIES OF SOVEREIGN FUNDS

Despite the increased use of the category of "sovereign investors" as one group, organizations in this groups are still very diverse. Therefore, we present below a set of sub-categories that have been used to group funds together. Our goal in presenting these sub-categories is to question to consistency of values and vocabulary among organizations in these sub-categories.

Justifying the category of sovereign investors based on traditional structural characteristics have been challenging. In 2008, Monk cited Rozanov noting that SWFs "…differ in size, age, structure, funding sources, governance, policy objectives, risk/return profiles, investment horizons, eligible asset classes and instruments, not to mention levels of transparency and accessibility" (2008).

In 2016, PwC underlined that SWFs were "a very diverse breed of heterogeneous institutional investors." However, extending the definition of sovereign investors to describe "all of the Government-related funds that are active in the global markets to achieve national objectives" did not change these categories. PwC also encouraged to "look at the Sovereign Investors' source of funding, entity status, age, objective, mandate and investment portfolio", almost the same categories as Rozanov.

In terms of larger categories, PwC notes the difference between Pension Funds and Sovereign Wealth Funds as two main subgroups of Sovereign Investors. A third group has been emerging in the academic literature: Sovereign Development Funds (Dixon & Monk,). Recently, the World Bank started a research project on this group, reinforcing the idea that it could become an increasingly popular category, even though few sovereign investors claim it as their identity (Halland, 2016).

In table 1, we list the criteria most commonly found in the literature to categorize and compare sovereign investors (PwC, 2016; Alsweilhem et al., 2016; Clark et al., 2013; Monk, 2009, Monk & Dixon, 2016). We propose to use some of these criteria as meta-data that could inform the grouping of sovereign investors, and some as key topics that could be discussed inside the annual reports. This informs our second hypothesis regarding sub-categories of sovereign investors.

Source of funding	
	Commodity vs non-commodity

	Taxes vs profit from government-owned companies
	Fiscal surplus, exchange-rate intervention based on reserve accumulation
Governance structure	
	Entity status (central bank, government-sponsored agency, independent public entity)
	Transparency & disclosure
	Internal governance and degree of political representation
	Investment process
Economic and Political Context	
	Age
	Political region vs dictatorship
Objective	
	Capital maximization
	Macro-economic stability
	Economic development
	Pension reserve in response to social welfare state crisis
Mandate	
	Domestic vs international
Investment portfolio	
	low-risk vs high-risk
	Exposure to real assets
	Passive vs active
	Size of Assets Under Management (AUM)
Region	
	Emerging vs developed economy
	5 continents
Mode of investment	
	Outsourcing vs Insourcing
	Co-investment platforms
Liabilities	
	Rule-based or discretionary