



HAL
open science

Governança dos oceanos: reconhecendo as instituições nomeadas em Direito internacional ambiental

Pierre Mazzega, Claire Lajaunie, Elisa Piccinini, Camille Rohrbacher,
Philippe Muller, Nathalie Aussenac-Gilles

► **To cite this version:**

Pierre Mazzega, Claire Lajaunie, Elisa Piccinini, Camille Rohrbacher, Philippe Muller, et al.. Governança dos oceanos: reconhecendo as instituições nomeadas em Direito internacional ambiental. A função do direito na gestão sustentável dos recursos minerais marinhos, Editora Processo - Rio de Janeiro, pp.75-106, 2021, 978-6589351498. hal-03099568

HAL Id: hal-03099568

<https://hal.science/hal-03099568v1>

Submitted on 6 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/340940517>

OCEAN GOVERNANCE: ON THE RECOGNITION OF INSTITUTIONS NAMED IN INTERNATIONAL ENVIRONMENTAL LAW

Chapter · September 2020

CITATIONS

0

READS

49

6 authors, including:



Pierre Mazzega

French National Centre for Scientific Research

142 PUBLICATIONS 1,588 CITATIONS

SEE PROFILE



Claire Lajaunie

French Institute of Health and Medical Research INSERM

74 PUBLICATIONS 194 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Future Health SEA [View project](#)



<https://revistas.ufg.br/bgg/article/view/44555> [View project](#)

Mazzege P., Lajaunie C., Piccinini E., Rohrbacher C., Müller P. and N. Aussenac (2020) Ocean Governance: on the Recognition of Institutions Named in International Environmental Law. In Oliveira C. Costa de, Lanfranchi M.-P., Barros-Platiau A. F., Galindo G. R. Bandeira (Orgs.) *A função do direito na gestão sustentável dos recursos minerais marinhos*. Rio de Janeiro: Editora Processo (no prelo).

OCEAN GOVERNANCE: ON THE RECOGNITION OF INSTITUTIONS NAMED IN INTERNATIONAL ENVIRONMENTAL LAW

Mazzege P.^a, Lajaunie C.^b, Piccinini E.^(c), Rohrbacher C.^(c),
Müller P.^(d) and N. Aussenac^(d)

^(a) CNRS, LISST UMR 5193, Laboratoire Interdisciplinaire Solidarités Sociétés Territoires, University Jean Jaurès, Toulouse, France

^(b) INSERM, LPED Laboratoire Population Environnement Développement, Aix-Marseille University, Marseille, France

^(c) Department of Language Sciences, University of Toulouse II, France

^(d) Institute of Research in Computer Science of Toulouse, University of Toulouse III, France

Abstract. This exploratory study serves two purposes: a) identify the institutions (organizations, programs, regulations, data infrastructures, etc.) named in some major International Environmental Law conventions and in the decisions or resolutions of the associated Conferences of the Parties (prior to 2015) and b) evaluate the potential contribution of Natural Language Processing - here SPACY and CoreNLP - in their Named Entity Recognition (NER) function, to this identification in an extensive body of legal texts.

After describing the main steps of computer-based NER and the performance criteria of the algorithms (recall and precision), we present the results of the analysis of about a thousand legal texts constituting our corpus. We combine the use of NER tools and the manual screening of raw results, a list of more than 800 institutions involved in environmental governance is established, 110 of them being more specifically involved in the governance of oceans and marine resources. The retrieval and classification of declarative information extracted from the corresponding websites then makes it possible to provide a picture of the governance of the oceans seen through these conventions, but also to specify the contributions and limits of the computer-based NER. The application of this method necessitates an interdisciplinary collaboration.

Keywords. Ocean governance, environmental law, natural language processing, text mining, named entities, governance modelling, institutions

1. INTRODUCTION

Presently the mapping of the actors, programs, initiatives, public policies, regulations, data infrastructures and other entities - that we group here under the term "institutions" - involved in the governance of the environment is not available. Whether it is on the environment and health issues, biodiversity conservation, oceans and marine resources management, production and dissemination of socio-environmental data, climate change mitigation measures, or on many other topics, the governance system, specific to each of these areas of activity is often mentioned but little known. If some entities are well identified - major international organizations, UN initiatives, international conventions, etc. - most of them are postulated but relegated to a *terra incognita*. Researchers, organizations, companies, decision makers, we all navigate by sight in the systems of environmental governance, without a map of any kind.

This fact has undesirable and, with a little hindsight, visible consequences. Thus, academic knowledge and discourse remain on an abstract level that while articulates important concepts, pays too little attention to the concrete facts shaping the complex and changing contexts of collective action. The potential for technological, institutional, managerial, training or collaboration innovation of the actors is hampered by the ignorance of the opportunities offered by the diversity of competences and the means existing to take in hand pressing problems. Large-scale architecture articulating decision-making mechanisms, distribution between actors of responsibilities, duties and obligations regarding environmental issues and sustainable development, is not drawn. In addition, many actions having impacts on the environment, natural resources, the living conditions of human or animal societies, biological, cultural and political diversities, remain outside the fields of analysis and evaluation based on evidence, and therefore out of publicity and opportunity to be debated even among key stakeholders. Moreover, these last two points raise delicate questions on the ethics of integrative research and of the political uses of scientific results, which are beginning to be taken in consideration and debated (Tuana, 2010; Resnik et al., 2016; Lajaunie and Mazzega, 2018).

The identification of the major institutions involved in environmental governance is important in view of a possible reversal of the persistent trend towards global environmental degradation (Ripple et al., 2017) and the massive and irreversible loss of biodiversity (Barnosky et al., 2011). Indeed, the future of our environment and our societies is a matter of real actors, governments and administrations, companies, non-governmental organizations, projects and collective initiatives, and digital entities (e.g. databases and knowledge, autonomous environmental sensors). The governance of the oceans and marine resources is becoming a major strategic issue (Kimball, 2001; European Commission - Consultation on international ocean governance¹, 2016), recognized internationally, and directly related to the sustainable development goal 14 (SDG14), as well as to other SDGs.

¹ "The aim of this consultation was to gather input on how the EU could contribute to achieving better international governance of oceans and seas to the benefit of sustainable blue growth." See https://ec.europa.eu/info/consultations/consultation-international-ocean-governance_en

The present study is based on the idea that the cartography of the ocean governance (and more broadly of the environmental governance) could be carried out on a large scale with the support of technologies related to the exploration of massive data sets produced by Environmental Law (Lajaunie et al., 2019). We thus pursue here two practical objectives:

- a. establish a list of institutions involved in ocean management based on a textual corpus of International Environmental Law conventions and decisions or resolutions of the associated Conferences of the Parties (COPs). In a later work, this list will feed the formal models of governance that we develop (Lajaunie and Mazzega, 2016a; Mazzega and Lajaunie, 2017; Mazzega et al., 2018) and which allow producing diagnoses on the governance of a domain and scenarios of changes;
- b. evaluate whether it is worth investing in the mastery of Natural Language Processing (NLP) software, here for the purpose of Named Entity Identification, or if developing a more limited and specialized approach to text mining is preferable. This second objective openly confers an exploratory character to this study. More investment would certainly give more complete and refined results (for example with a preliminary "training" of the algorithms) but would induce the adoption of the tools that we wish to evaluate.

Section 2 presents the basics of detecting named entities by text mining, and the problem of categorizing these entities. Section 3 describes the text corpus used and the method used to detect named entities. The results are summarized in Section 4 and evaluated for their representativeness and relevance to feed governance models. The perspectives drawn by this study and our conclusions are gathered in Section 5.

2. TEXT MINING AND RECOGNITION OF NAMED ENTITIES

2.1 COMPUTER-ASSISTED RECOGNITION OF NAMED ENTITIES

The analysis of large textual corpora is now based on tools from Natural Language Processing (NLP; Feldman and Sanger, 2007; Berry and Castellanos, 2008). The role of human agents in mining text - e.g. elaboration or choice of a corpus, supervision of the learning phases of the algorithms, screening and interpretation of results, expertise - is more or less important according to the origin and type of texts considered (e.g. html pages, PDF files) and according to the technical options embedded in the software executed by a machine. The major functions of these approaches concern the extraction of events, entities or relations between entities (e.g. in the field of legal studies: Peters and Wyner, 2016; Lajaunie and Mazzega, 2016b; Lajaunie et al., 2018), sentiment analysis and opinion mining, the categorization of texts, topic modelling, thesaurus building, etc. Their field of application covers all sectors using information in the form of texts available in natural language (digitized archives, web pages, technical documentation, etc.).

Named Entity Recognition and Classification (NERC) is a special case of recognizing and extracting information from a textual corpus (Nadeau and Sekine, 2007). It consists in identifying common names or proper names in the texts and storing them in predefined classes such as "names of people", "names of organizations", "names of places", electronic or geographical "addresses", "bibliographic references", "hyperlinks", etc. The chosen

classes depend on the software used or on the thematic application in a phase of post-processing of the results.

Numerous works are currently underway to produce powerful NER or NERC tools. Dlugolinský et al. (2013) present the different modes of operation of six Named Entity Recognition Software. They use a rather small corpus composed of micro-posts² without standardized punctuation. They annotate it by hand with the thematic groups expected in order to have a corpus available for testing the six NER tools. Jiand et al. (2016) also present an evaluation of different Named Entity Recognition systems and combine several systems with each other. They explain that the heterogeneity of domains, methodologies and languages does not make it easy to choose the right NER parameters according to the body of work. In addition, the different types of entities and media in input or of formats in output further complicate the choice of parameters for the user. Finally, they build a hybrid NER tool that combines all the useful parameters for their area of interest. In fact, software performance depends both on the underlying linguistic theory instantiated in operational analysis tools, and on the algorithmic choices made.

The main problems related to the detection of named entities are the orthographic variation, the ambiguity of certain terms and the polysemy. The anticipation of difficulties related to the literary style, gender, but also punctuation, the consistency of the use of capital letters (important point for the detection of named entities), spelling errors, word spacing, or text formats is not obvious. In addition, as previously stated, the NER is used in various fields of application which constitute new constraints to be taken into account. As it is well known, the field of specialty (e.g. medicine) can strongly influence the way NERC is conducted. Indeed, the use of a specialized vocabulary constitutes a case of analysis far removed from the kind of textual corpus with which the system of detection is usually trained. (Kanimozhi, 2017; Song et al., 2015). This type of difficulty is partly overcome thanks to the use of ontologies or specific dictionaries or even a preliminary phase of manual annotation of texts. Moreover, the standard procedure for evaluating the performance of a NER or NERC tool consists in comparing the results obtained by this tool with the exhaustive list of named and classified entities obtained by a "manual" analysis of a test corpus. Incidentally, this means that the present study can in no way be used to evaluate the performance of the tools we use, because that would imply that we first do the desired detections by hand (which is not our goal).

2.2 PRECISION AND RECALL

Two concepts are central to evaluating the performance of a NERC tool: precision and recall (see Box 1). A 100% precision means that all entity detections are exact. However, this does not imply that the detection is exhaustive, that is to say that certain entities have not been forgotten. A 100% recall indicates that all the entities present in the textual corpus have been identified. However, the recognition can be understood in a strict sense - the

² A micro-post is a very short message posted on the internet, not exceeding 140 to 200 characters in general.

detected entity corresponds exactly to the entity named in the text (with capital letters and exact spelling) - or with a certain flexibility, for example a detected entity reproducing only a part of a named entity being considered a positive detection. Here we use the flexible approach: for example, we consider that the "Arctic Council's Conservation of Arctic" entity produced by the software is an acceptable detection of the "Arctic Council's Conservation of Arctic Flora and Fauna", a working group of the Arctic Council. It should be noted, however, that this decision is *ad hoc*, since this specific detected entity unambiguously identifies this working group, which can be only checked *a posteriori* and manually.

Box 1. Performance of Named Entity Recognition Tools: Precision and Recall

When a named entity recognition is carried out, three main cases are to be considered:

- True Positive: an entity is correctly recognized;
- False Positive: the recognition did not have to be;
- False Negative: a recognition that should have occurred was not made.

Let TP, FP and FN be the number of true positives, false positives and false negatives respectively taking place during the analysis of a textual corpus. The precision P and the recall R are respectively defined (in %) by:

$$P = 100 \times TP / (TP + FP)$$

$$R = 100 \times TP / (TP + FN)$$

A 100% precision is obtained when there is no false positive (and $TP \neq 0$). The recall is 100% when there is no false negative (and $TP \neq 0$). Accuracy and recall are zero when there is no true positive ($TP=0$).

In this study we use two software: SpaCy and CoreNLP. SpaCy³ is a free open-source library for NLP in Python. It is very practical and easy to use. It uses vector representations for words, which is a proven method in NLP and widely used in information retrieval and detection (see e.g. Mikolov et al., 2013 in relation to large data sets). SpaCy is known in the NLP community for good recall performance but with low precision. We also use CoreNLP⁴ (Manning et al., 2014), mainly in chain, after SpaCy, to try to correct the low precision of the latter while enjoying its good recall rate. In addition, CoreNLP is able to display the co-references⁵, unlike SpaCy. Chang and Manning (2014) propose a parameter ("TokensRegex") of the CoreNLP software which allows to define rules in cascade using regular expressions. The combined use of the NER of the software and of the TokensRegex makes it possible to achieve more efficient named entity recognition because the extraction of information or matching from tokens is simplified. In addition, Angeli et al. (2013) show that CoreNLP is a powerful and reliable system for information retrieval. It is therefore relevant to use it for a NERC task.

³ Voir le site <https://spacy.io/> (Accessed 5 Feb 2019)

⁴ Voir le site <https://stanfordnlp.github.io/CoreNLP/index.html> (Accessed 5 Feb 2019)

⁵ The fact that two expressions or sentences have the same referent.

2.3 INSTITUTIONS' MAIN TYPES AND DOMAINS

The screening of the list of named entities identified by text mining (see Sec.3 below) led us to carry out a first filtering and to define as an analytical framework the main types and areas of competence of these entities. We are now introducing the approach adopted to understand the nature and scope of the results presented in the following sections. The proposed examples are all drawn from the raw list of named entities obtained in this study.

First of all, we only selected singular entities. For example, we do not use the term "regional fisheries management organization" (RFMO) or "non-governmental organization" (NGO) because they are classes. But we retain any particular RFMO or NGO quoted by name. The concepts - such as "exclusive economic zone" (EEZ), "European marine site" (EMS), "epistemic community", "ecological network guidance" - are not selected. They are detected as named entities probably because of their use with uppercase initials, and their grammatical function within sentences (noun phrase).

In the expressions "secretariat of the Convention *lambda*" we have kept only the identification of the convention (say, "Convention *lambda* "). Sometimes we have also reduced the number of entities selected, when the name of the parent institution largely and unambiguously covers that of daughter institutions for which it was difficult to collect complete information. Thus, from the three entities named "International Coral Reef Initiative", "International Coral Reef Action Network" and "International Coral Reef Information Network", we retained only the first. The components of the Antarctic Treaty System have not been conserved, the area of jurisdiction being primarily continental. The concept of "Party" to a conference is discarded (as a class) as well as the name of countries or governments, ocean governance being only a small part of their activities.

We could continue the description of the cases and decisions made. But these remarks aim above all at showing that the choice of named entities seems difficult to establish on the basis of objective, simple or automatable criteria. It closely depends on the objectives of the study and the culture of the analyst. The same comment applies to the choice of typology of named entities and the domains in which they operate.

The main types of named entities used for the posterior analysis of results are: private company; data infrastructure (repository, database, data or knowledge system); network, alliance, partnership or consortium; legal norm, policy or regulation; organization (non-profit, NGO, intergovernmental, etc.); program, initiative or plan; geographic site (e.g. park).

Each singular entity can be involved in one or several domains from the following list:

- Bioprospecting (of marine living resources);
- Capacity building, training;
- Certification, standardization;
- Climate change;
- Conservation;
- Cooperation;
- Development (in particular in relation with international goals or targets);

- Energy from the ocean (offshore wind, marine current or tidal power, ocean thermal energy);
- Fisheries or aquaculture;
- Food production, food security;
- Heritage;
- Management of marine resources, area, policy implementation,...;
- Mining or extraction of marine minerals, fossil fuels, materials;
- Monitoring and reporting, particularly in relation to compliance with legal obligations or norms;
- Navigation or shipping;
- Policy, law or regulation advising, design or implementation;
- Pollution;
- Research, observation, data acquisition;
- Security and prevention in the maritime domain;
- Tourism;
- Trade.

The relevance of associating a domain with an entity is assessed with regard to the declarative information available on the entity's website. We subjectively evaluate whether each area is at the heart of the jurisdiction or - if applicable - of the entity's mission or activity. For example, most organizations will report "cooperation" without the implementation or facilitation of cooperation being one of their missions or an identity trait (in which case this area is not associated). The "monitoring" activity is distinguished from that of the observation made by scientists of marine environments and resources in that it aims at producing information enabling to assess the normative commitments or realization of environmental or development objectives or targets. Of course, all the domains in the list concern the oceans even when this is not specified in our list of domains (e.g. "heritage" or "pollution").

Because of the nature of the material analyzed (information provided in natural language), obviously the results of the analysis are not to be interpreted strictly or quantitatively. Nevertheless, they provide clear and informative insights into ocean governance as seen through major conventions of International Environmental Law.

3. THE CORPUS OF LEGAL TEXTS AND ITS PROCESSING

3.1 PREPROCESSING OF LEGAL TEXT CORPUS

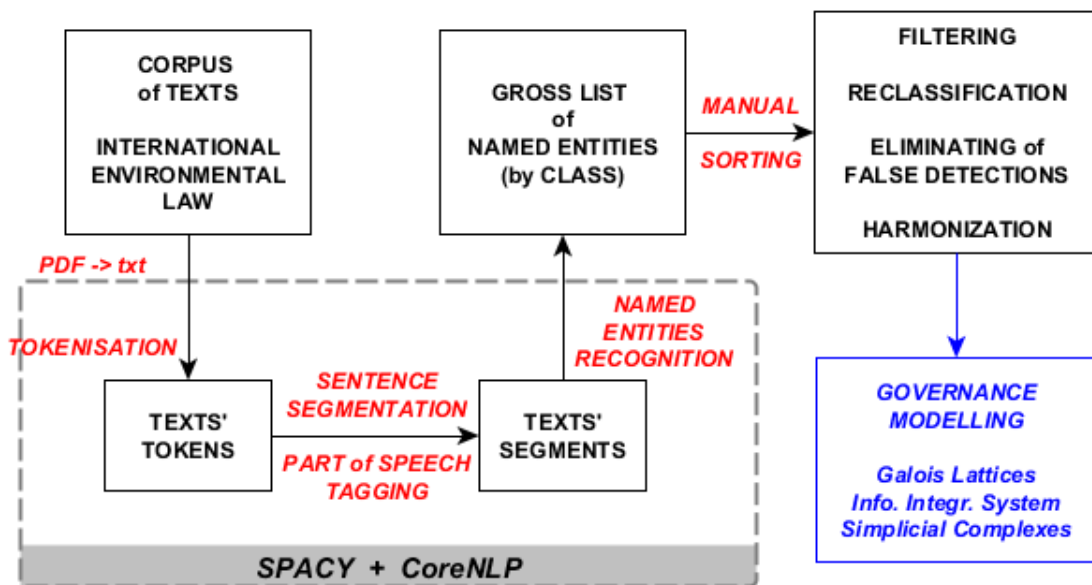
Our corpus contains about 1,000 texts corresponding to the following conventions and the decisions or resolutions adopted by the Conferences of the Parties (COPs) until the end of 2014:

- CBD: Convention on Biological Diversity, 1992;
- CITES: Convention on International Trade in Endangered Species of Wild Fauna and Flora, 1973;
- CMS: Convention on the Conservation of Migratory Species of Wild Animals, 1979;

- UNFCCC: United Nations Framework Convention on Climate Change, 1992;
- Kyoto Protocol to the UNFCCC, 1997.

All texts were downloaded from convention websites, most of them in PDF format, and converted to "text" files (Uniform Type Identifier: public.plain text; with ASCII character set). The files corresponding to each COP of the UNFCCC have been split so that each decision corresponds to a file. No further pre-processing of the documents has been done, in particular to eliminate any slag produced during the conversion of PDF to text format, leaving the software with the task of finding the grammatical structures well formed in the analyzed files.

Figure 1. Overall Information flow and processing associated with the Named Entities Recognition.



The main steps of the analysis are summarized in the flow diagram of Figure 1. The successive phases of tokenization, sentence segmentation, part of speech tagging and named entities recognition are provided by SPACY or by CoreNLP (see details below). The raw list of identified entities is retrieved at the end of the software processing and screened by hand in order to mainly perform the filtering, and when necessary, the re-categorization of each entity. We then visited the websites associated with each of the selected institutions to determine their type and areas of expertise or activity.

3.2 NAMED ENTITIES RECOGNITION

A first analysis of the whole corpus is carried out with SPACY in order to find the entities that the software classifies under the categories "country, city, State", "organization", "person", "event", "law", or "word of art", the other categories⁶ being neglected. SPACY

⁶ See <https://spacy.io/api/annotation#named-entities>

returns a raw list of about 8,000 entities to which we add nearly 10,000 organization names from a Wikipedia list⁷.

In a second step, we use the NERC function of CoreNLP, in order to extend this list once again, while benefiting from the more performant precision of CoreNLP. Recognition of Named Entities of CoreNLP, like that of SpaCy, takes some parameters into account (capital letters, punctuation, syntactic-lexical context, etc.) and then classifies the different Named Entities according to a different categorization than in SPACY, namely according to the thematic groups "organization", "miscellanies", "location" and "event".

In order to add SPACY thematic group annotations to CoreNLP, we use CoreNLP's *TokensRegex*⁸ feature. This parameter makes it possible to detect additional Named Entities, here those identified by SPACY, thanks to the rules that we transmit to it. To do this, a correspondence is established between the categories assigned by SPACY and the CoreNLP categories. The analysis was also programmed to keep track of the identity of the text in which a named entity appears as well as the appearance context - mainly the relevant sentences. This information - which we do not exploit here - can be used to carry out targeted post-processing of sentences (for example as the matching of nominal groups) or to associate a lexical context of use of each of the named entities.

Table 1. *Extracted from the raw list of named entities detected by SPACY and CoreNLP. The second column indicates whether an entity is retained ("Y") after manual filtering (otherwise "N"). The last column defines the named entity.*

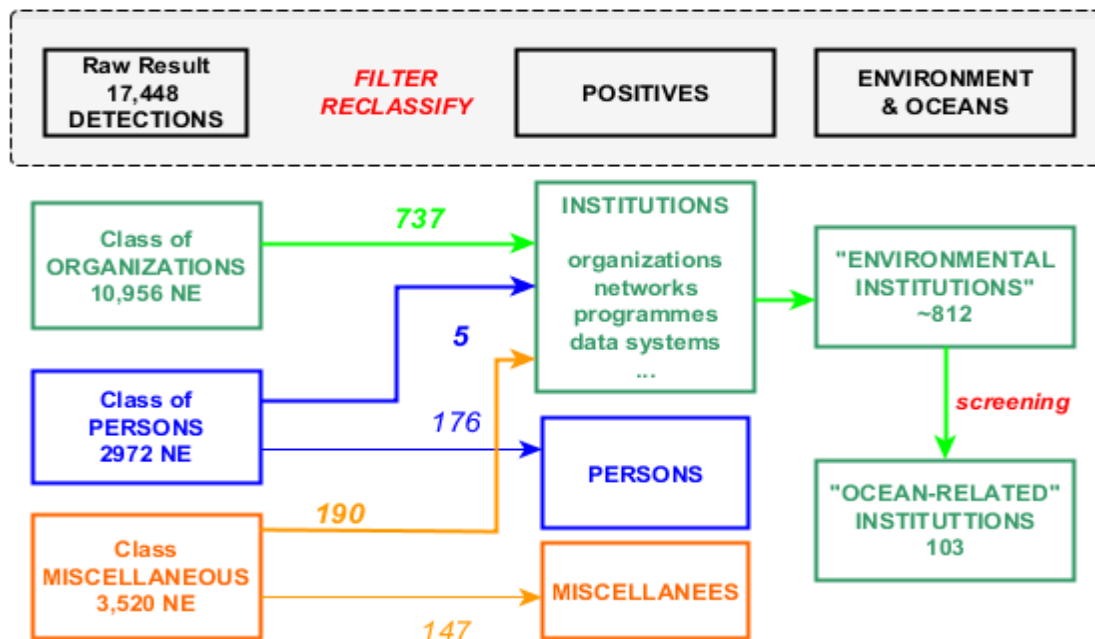
Named Entity		Basic definition
...		
Uroplatus	N	Kind of gecko of the Gekkonidae family
D. Participation	N	Start of list paragraph
BVC Holding S.A.	Y	Company
MS Word	N	Software
ESTABLISHES	N	Verb written in capital letters
CITES Management	N	Expression (involving the CITES)
United Nations Environment Programme	Y	Programme
Target	N	Common name (or part of expression such as "Aichi target")
Gekkonidae	N	Gecko family
FAO Global Plan of Action for Plant Genetic Resources for Food	Y	(Partial) name of a FAO plan of action.
UNEA	Y	Acronym of the UN Environment Assembly
Warsaw	N	City
the UK Government	N	Government
Saker Falcon Network	Y	Scientific network
...		

⁷ https://en.wikipedia.org/wiki/List_of_environmental_organizations

⁸ <https://stanfordnlp.github.io/CoreNLP/regexner.html>

We are only using the list of named entities here. Out of SPACY and CoreNLP applied to the complete corpus, we have a raw list of more than 17,000 named entities. The results are very noisy as shown in Table 1. There are also multiple detections of the same entity (e.g. truncated name). The number of named entities in the raw list by categories, and then their number once filtered and redistributed, are shown in the diagram in Figure 2. Once hand-screened, there are only about 800 institutions linked to the environmental governance, and of these, about 110 are involved in ocean governance (see the list in the appendix).

Figure 2. Filtering and re-categorization of the 17,448 Named Entities of the raw list obtained with SPACY and CoreNLP. The number of named entities that are retained and moved from one category to another is indicated on each link between categories. The centre column only lists the true positives. Thus, the category "institutions" (top of the central column) is made up of 737 entities coming from the (SPACY + CoreNLP) category "organizations", 5 which had been initially identified by the algorithms as "persons" and 190 as "miscellanies".



4. ANALYSIS AND EVALUATION OF RESULTS

The identification of named institutions is only the first step towards modelling ocean governance. The models we develop (Lajaunie and Mazzega, 2016a; Mazzega and Lajaunie, 2017; Mazzega et al., 2018) use other information such as the description of the missions or competences of the organizations, the major themes of the conventions, regulations or political texts, the composition of the governing boards or if any, the training offers of the organizations. This information is drawn from the institutions' websites, here the 110 institutions identified as participating in the ocean governance. In order to give a kind of profile of this governance seen through the CBD, CMS, CITES and UNFCCC conventions, we produce basic statistical elements on the types and competences of these institutions.

4.1 DISTRIBUTION OF TYPES OF NAMED INSTITUTIONS

Figure 3 shows the number of institutions identified for each of the types used (*a posteriori* typology). Each institution is associated with only one type. Organizations - non-profit, governmental, non-governmental, ... - form the most represented type, with 49 registered organizations. The second most represented type is "programs, initiatives, plans" with 18 entities.

The most striking feature is probably the absence of companies named in the conventions. As we shall see, organizations are particularly involved in environmental or biodiversity conservation, or in monitoring the effects of actions taken to preserve the marine environment or to ensure sustainable use of marine resources. But companies whose activities rely on the exploitation or use of maritime areas or resources or have significant impacts on the marine environment and biology, are not specifically named.

In 2000, the United Nations decided to develop a common framework for UN-Business collaboration, the Guidelines on a Principle-based Approach to the Cooperation between the United Nations and the Business Sector⁹ which apply to the UN Secretariat as well as separately administered organs, funds and programmes.

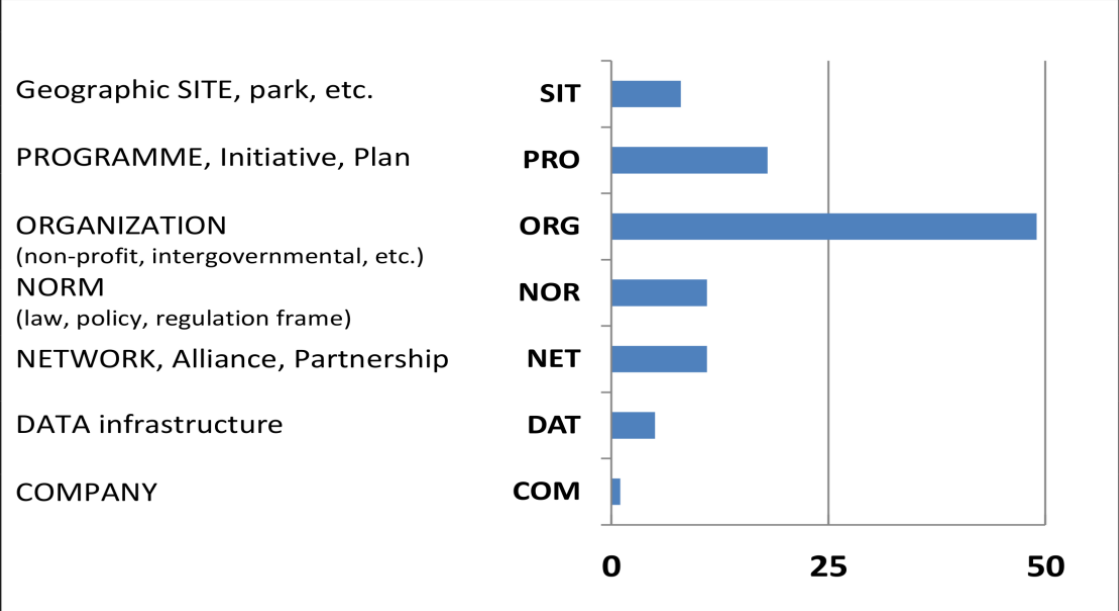
The role of the private sector into the realization of the environmental agreements' goals has also been underlined on several occasions by the various conventions examined. The COP 11 of the UNFCCC, by the decision 2/CP.11, created a programme of work of the Subsidiary Body for Scientific and Technological Advice (SBSTA) to assist all Parties, in particular developing countries, "to improve their understanding and assessment of impacts, vulnerability and adaptation, and to make informed decisions on practical adaptation actions and measures to respond to climate change on a sound, scientific, technical and socio-economic basis". This programme became the "Nairobi work programme on impacts, vulnerability and adaptation to climate change" (COP 12, Nairobi, Statement by the Executive Secretary). This programme created the Private Sector Initiative (PSI) "to catalyze the involvement of the private sector in the wider adaptation community" and launched in 2012, the PSI Database showing case studies and good practices and climate change adaptation activities undertaken by the private sector. The companies are named in this database but are not included in the decision process of the UNFCCC and thus not identifiable as such in the corpus we examined.

The involvement of private companies into the CBD process dates back to the COP 3 Decision III/6 (1996) which urged the executive secretary to explore possibilities for encouraging the private sector in supporting the Convention's objectives. The need to involve the private sector has been reaffirmed into various following decisions of the COPs to the CBD (such as Decision V/11; Decision VI/26, Annex Strategic Plan for the Convention on Biological Diversity 2002-2010 Goal 4.4 or Decision IX/26). The UN Guidelines have been reviewed and updated in 2009 (UN General Assembly resolution 68/234) highlighting the increased recognition by Member States that the business sector could play an important role in support of the 2030 Agenda for Sustainable Development. The Guidelines state that

⁹ See <https://business.un.org/en/documents/5292> (accessed 22 Feb 2019)

cooperation between the United Nations and the business sector is based on principles relating to human rights, labour, the environment and anti-corruption integrated in intergovernmental agreements, and that are specifically relevant for business (United Nations, 2009).

Figure 3. Number of institutions (named in the corpus and identified by NERC) by type.



In line with these guidelines, the CMS developed a code of conduct for partnerships with the private sector relying on the same principles as those affirmed by the United Nations (integrity, transparency, commitment, impartiality...). This code aims at “the improvement of the environmental impact of the private sector, awareness raising of the value of migratory species and the creation of support for the conservation of migratory species through increased local, national and regional investments” (CMS, 2009).

Regarding the CITES, various initiatives to raise awareness about endangered species have been decided whether they concern the creation, in collaboration with a group of entrepreneurs, of a smartphone application tracing the origin of a wildlife product (application Asking, Nest’up, 2013) or the sector of innovation more broadly (World Economic Forum, 2013). The work of the Standing Committee and of the COP focuses on trade and on the implementation of compliance measures and thus refer to the States and the measures they adopt and not to the companies namely.

In 2010, the decision X/21 on Business Engagement of the COP 10 of the CBD insisted on the necessity to encourage the establishment of the national and regional business and biodiversity initiatives by facilitating a forum of dialogue among Parties and other Governments, business, and other stakeholders; to compile information on existing tools able to further facilitate the engagement of businesses in integrating biodiversity concerns into corporate strategies and decision-making and to encourage the development and application of such tools and mechanisms.

It has been followed by the organisation in 2011 of the First meeting of the Global Platform for Business and Biodiversity aiming at promoting markets that support nature conservation and sustainable use and that incorporate the value of biodiversity in business operations. Since then, such a meeting is organized every year and thanks to the activities of the Joint Liaison Group of the various conventions related to biodiversity (and among them CBD, CMS and CITES), it benefits to the whole group of conventions.

Nevertheless, these activities are not included in the decisions or resolutions of the COPs and as for the UNFCCC, the name of companies from the private sector do not appear in the body of decisions of the COPs and are out of the corpus we considered (even before filtering the raw results for more specific ocean issues).

Data infrastructures - data repositories, databases, data or knowledge system - are also very little mentioned in the conventions and COPs decisions or resolutions. This may seem surprising, the demand for data, observations, indicators, becoming a leitmotif of various texts of strategy or policies at international and regional level (see e. g. UNSG-IEAG, 2015; European Commission, 2015; Mazzucato, 2018). In particular the adoption of the marine spatial planning approach and its implantations on many sites in the world, leads to use a broad spectrum of physical, biological but also socio-ecological (MSP Data Study, 2016) data.

The low occurrence of data infrastructure designations in our corpus is explained by the fact that these infrastructures generally cover an area that cannot be related to the oceans alone. For example, databases on biodiversity concern a whole phylum or class of life. Physics databases often include several large "compartments" of the environment, but possibly name the ocean as one of them (i. e. the National Oceanic and Atmospheric Administration). On the other hand, the names of indexes or indicators generally explicitly indicate the medium to which they relate (e. g. Marine Trophic Index, Ocean Biogeographic Information System). In the end, data infrastructures are named in more specialized texts.

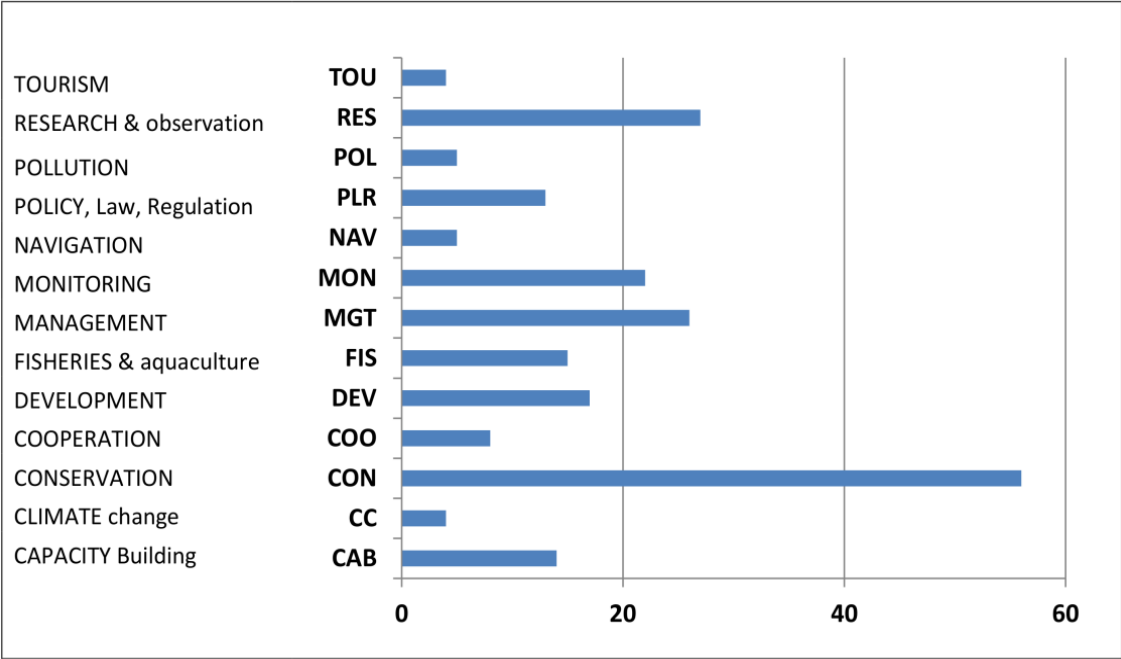
4.2 DISTRIBUTION OF DOMAINS OF NAMED INSTITUTIONS

Each institution can operate in one or more of the domains listed in Figure 4. Eight domains of our initial list are not represented, being associated with fewer than 4 institutions: bioprospecting; certification; marine energy; food; heritage; mining; security; trade. Of course, FAO focuses on food security and production, and the International Seabed Authority on marine mining in the international seabed area beyond the limits of national jurisdiction. But we did not find any other organizations, and no companies, whose activities would be directed - even partially - towards one of these domains. Bio-prospecting, certification and energy production are mainly operated by companies. The regulation of these activities via the conventions and COPs does not namely mention these companies. Trade is likely covered by the "navigation and shipping" domain.

Our results (Figure 4) suggest that the governance of the oceans is seen by the CBD, CITES, CMS and UNFCCC conventions mainly from the point of view of conservation. Management and also monitoring activities with regard to the commitments contracted by States are important, with the support of scientific research. Activities related to sustainable development (especially in connection with Sustainable Development Goals, and previously

the Millennium Development Goals), fisheries and aquaculture regulation, and capacity building, come second.

Figure 4. Number of institutions (named in the corpus and identified by NERC) by domain of activity.



Although the role of the oceans in climate and global warming is well known and is the subject of much research, this link in terms of environmental governance does not appear in the conventions. As we have already seen above, if the private sector or business sector is mentioned generically in the decisions of the COPs of the different conventions studied, the names of specific company do not appear. Furthermore, organizations working on the link between climate change and other environmental changes have *de facto* a broad spectrum of study or of activity and they are not limited to a specific topic such as the ocean. Thus, when looking specifically on names entities in relation to the ocean it is not surprising to find only few of them having an interest in climate change.

Incentive and support for cooperation, at the global, regional or interstate level, tend to remain the preserve of major international organizations or major programs, including the United Nations.

4.3 SOME PERSPECTIVES

This exploratory work opens various perspectives. First, we will establish similar results based on the more than 800 named institutions already identified and involved in environmental governance. The exploration of as many websites, while tedious, is manually feasible, but would be greatly facilitated by an automated site visit and the use of NERC tools or text mining applied to the content of these sites. All this information can be ordered and analyzed in the form of various networks (e.g. information networks, Galois lattices, simplicial complexes) corresponding to as many aspects of the ocean or environmental

governance. Of course, the textual corpus can be expanded to include other conventions (e. g. Ramsar Convention on Wetlands of International Importance) or related *ad hoc* scientific group publications. In all cases our corpus must be incremented with the COPs' decisions and resolutions published after 2014.

It will also be interesting to compare the governance maps we will produce with the analytical framework used by various academic authors in their ocean governance monographs (eg following specific angles of view like marine protected area in Caveen et al. , 2015 or international environmental law in Zacharias, 2014) or the environment, either by large international organizations through texts that they produce and disseminate widely (e.g. UNEP, 2016).

Finally, in a broader perspective of knowledge mapping and conceptual modelling of the multi-scale governance of the oceans and marine resources, the exploitation of the lexical context of the occurrence of the named entities should provide essential information for the understanding of such complex and large-scale governance systems.

5. CONCLUSION

The first objective of this exploratory study was to evaluate whether it is justified to persevere in the mastery of Natural Language Processing tools, here in their functionality of Named Entity Recognition. The NLP surely provides a much more contextual, integrative and broad-based view of international environmental law, notably by analyzing important textual corpora. The availability of large sections of law, international or national, in digital form, opens the prospect of Big Data Analytics approaches, which lawyers could usefully take hold of. However, it is preferable to carry out this type of analysis in the framework of interdisciplinary collaborations, thus ensuring not only the mastery of the legal subjects, but also of the linguistic analysis and of computer-based approaches.

The mapping of the institutions involved into the governance of the oceans (or more broadly, of the environment), from the global to the local level is now possible thanks to these approaches and must make it possible to carry out diagnostic analyzes (allowing to respond to questions such as: how governance could be improved to be more effective with regard to the stated objectives?) and scenarios (what consequences can be expected from changes in the context of multi-level cooperation and regulation?) of governance. After the identification of the institutions named in the conventions and COPs decisions or resolutions (prior to 2015) - here 110 institutions involved in the governance of oceans and marine resources (list in the Appendix), drawn from a list of more than 800 institutions identified in the CBD, CMS, CITES and UNFCCC conventions - NLP tools should be used to explore the contents of institutions' websites and draw information about the activity or areas of competence of these institutions, their composition and possible inter-linkages.

Computer-based analysis allows mapping a myriad of various institutions - i. e. organizations, programs, legal norms, public policies, data infrastructure ... - and helps having an comprehensive overview of the governance system that is impossible to get otherwise.

Acknowledgements. PM and CL are grateful to Prof. Carina Costa de Oliveira (Faculty of Law, University of Brasilia) for the invitation to participate in the workshop "Sustainable Management of Marine Mineral Resources: Cross Perspectives of International and National Rights (Brazil and France)", 4-5 June 2018, University of Brasilia, Brazil where this work has been presented for the first time. This work was presented at the 2018 IUCN Academy of Environmental Law Colloquium "The Transformation of Environmental Law and Governance: Innovation, Risk and Resilience", 4-6 July 2018, University of Strathclyde, Glasgow UK.

6. REFERENCES

- Angeli G., Chaganty A. T., Chang A. X., Reschke K., Tibshirani J., Wu J. et al. and C. D. Manning (2013) Stanford's 2013 KBP System. In TAC.
- Barnosky A. D., Nicholas M., Tomiya S., Wogan G. O., Swartz B., Quental T. B., Marshall C., McGuire J. L., Lindsey E. L., Maguire K. C., Mersey B. and E. A. Ferrer (2011) Has the Earth's sixth mass extinction already arrived? *Nature*, 3 March 2011, 471: 51-57. doi:10.1038/nature09678
- Berry M. W. and M. Castellanos (eds.) (2008) *Survey of Text Mining II: Clustering, Classification, and Retrieval*. Springer-Verlag London. DOI 10.1007/978-1-84800-046-9
- Caveen A., Polunon N., Gray T.a and S. M. Stead (2015) The controversy over marine protected areas. *Science meets policy*. Springer Briefs in Env. Sci., Springer, Switzerland, xix+162 pp.
- Chang A. X. and C. D. Manning (2014) TokensRegex: Defining cascaded regular expressions over tokens. Tech. Rep. CSTR 2014-02.
- CMS (2009) Code of conduct for partnerships with the private sector. Convention on the Conservation of Migratory Species of Wild Animals CMS/StC36/16 Agenda Item 9.d https://www.cms.int/sites/default/files/document/Doc_16_code_of_conduct_0.pdf (Accessed 22 Feb 2019)
- Dlugolinský Š., Ciglan M. and M. Laclavík (2013, June) Evaluation of named entity recognition tools on microposts. In: 2013 IEEE 17th International Conference on Intelligent Engineering Systems (INES), 197-202.
- European Commission (2015) Research Call Horizon 2020, Policy-development in the age of big data: data-driven policy-making, policy-modelling and policy-implementation, [H2020-SC6-CO-CREATION-2016-2017](https://ec.europa.eu/info/consultations/consultation-international-ocean-governance_en)
- European Commission - Consultation on international ocean governance (2016) Summary of the results of the public consultation on international Ocean Governance. See: https://ec.europa.eu/info/consultations/consultation-international-ocean-governance_en (Accessed 24 Feb 2019)
- Feldman, R. and J. Sanger (2007) *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge: Cambridge University Press.

- Jiang R., Banchs R. E. and H. Li (2016) Evaluating and combining named entity recognition systems. In Proceedings of the Sixth Named Entity Workshop, Joint with 54th Association for Computational Linguistics, 21-27.
- Kanimozhi U. (2017) A bootstrapping approach for entity linking from biomedical literature. *Asian Journal of Pharmaceutics*, 11(1). DOI: <http://dx.doi.org/10.22377/ajp.v11i01.1106>
Available on: <https://www.asiapharmaceutics.info/index.php/ajp/article/view/1106/618>
Accessed 9 April 2018
- Kimball L. A. (2001) *International Ocean Governance: using International Law and organizations to manage marine resources sustainably*. IUCN, Gland, Switzerland and Cambridge, UK. xii + 124 pp.
- Lajaunie C., Schafer B. and P. Mazzega (2019) Big Data x Environmental Law: massive data technology enters environmental law. *Submitted*.
- Lajaunie C. and P. Mazzega (2018) A pragmatic approach of ethics in interdisciplinary research on biodiversity conservation. *Spec. Issue on Ethics of Biodiversity Conservation, Asian Bioethics Review* 10(4):241-243.
- Lajaunie C., Mazzega P. and R. Boulet (2018) Health in biodiversity-related conventions: analysis of a multiplex terminological network (1973-2016). In: S.-H. Chen (ed.), *Big Data in Computational Social Science and Humanities, Computational Social Sciences*, p. 165-182. https://doi.org/10.1007/978-3-319-95465-3_7
- Lajaunie C. and P. Mazzega (2016a) Organization networks as information integration system - Case study on environment and health in Southeast Asia. *Advances in Computer Science: an International Journal*, Vol. 5, 2(20), 28-39 <http://www.acsij.org/acsij/article/view/461>
- Lajaunie C. and P. Mazzega (2016b) Mining CBD. *Brazilian Journal of International Law*, vol.13(2), 277-292. doi: 10.5102/rdi.v13i2.4058.
- Manning C. D., Surdeanu M., Bauer J., Finkel J., Bethard S. J. and D. McClosky (2014) The Stanford CoreNLP natural language processing toolkit. In Proc. 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, 55-60.
- Mazzega P., Lajaunie C. and E. Fieux (2018) Governance modelling: dimensionality and conjugacy. In *Graph Theory - Advanced Algorithms and Applications*, Beril Sirmacek (Ed.), ISBN 978-953-51-3773-3, InTech Publisher, Reijika, Croatia, 63-82. <http://dx.doi.org/10.5772/intechopen.71774>
- Mazzega P. and C. Lajaunie (2017) Modelling Organisation Networks Collaborating on Health and Environment within ASEAN. In *Complex Systems: Theory and Applications*, R. S. Martinez (Ed.), NOVA Publ. Hauppauge, NY- USA, ISBN: 978-1-53610-871-2, Chap. 5, 117-148.
- Mazzucato M. (2018) *Mission-Oriented Research & Innovation in the European Union, A problem-solving approach to fuel innovation-led growth*, European Commission, Directorate-General for Research and Innovation.
- Mikolov T., Chen K., Corrado G. and J. Dean (2013) Efficient estimation of word representations in vector space. Available on <https://arxiv.org/pdf/1301.3781.pdf> (Accessed 22 Feb 2019)

- MSP Data Study (2016) MSP Data Study Executive Summary. Technical Study under the Assistance Mechanism for the Implementation of Maritime Spatial Planning. European Commission - Directorate-General for Maritime Affairs and Fisheries, Brussels, 131 pp. DOI:10.2826/25289
- Nadeau D. and S. Sekine (2007) A survey of named entity recognition and classification. *Linguisticae Investigationes* 30(1): 3-26.
- Peters W. and A. Wyner (2016) Legal Text Interpretation: Identifying Hohfeldian Relations from Text. In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. http://www.lrec-conf.org/proceedings/lrec2016/pdf/253_Paper.pdf (Accessed 22 Feb 2019)
- Resnik D. B., and K. C. Elliott (2016) The ethical challenges of socially responsible science. *Accountability in Research* 23 (1): 31–46. <https://doi.org/10.1080/08989621.2014.1002608>
- Ripple W. J., Wolf C., Newsome T. M., Mauro Galetti M., Alamgir M., Crist E., Mahmoud M. I., Laurance W. F. and 15,364 scientist signatories from 184 countries (2017) World scientists' warning to humanity: a second notice. *BioScience* 67(12):1026–1028, <https://doi.org/10.1093/biosci/bix125>
- Song M., Kim W. C., Lee D., Heo G. E. and K. Y. Kang (2015) PKDE4J: Entity and relation extraction for public knowledge discovery. *Journal of biomedical informatics* 57:320-332.
- Tuana N. (2010) Leading with ethics, aiming for policy: new opportunities for philosophy of science. *Synthese* 177 (3): 471–492. <https://doi.org/10.1007/s11229-010-9793-4>.
- UNEP (2016) Regional oceans governance. Making regional seas programmes, regional fishery bodies and large marine ecosystem mechanisms work better together. Billé R., Lucien Chabason L., Petra Drankier P., Erik J. Molenaar E. J., Julien Rochette J. (authors). *UNEP Regional Seas Reports and Studies No. 197*, xvi+212 pp.
- United Nations (2009) Guidelines on a principle-based approach to the Cooperation between the United Nations and the business sector, 9 p. https://www.unglobalcompact.org/docs/news_events/9.1_news_archives/2009_11_23/un_business_guidelines.pdf (Accessed 22 Feb 2019)
- UNSG-IEAG (2015) United Nations Secretary-General's Independent Expert Advisory Group, *A World That Counts: Mobilising the Data Revolution for Sustainable Development*, United Nations, 2015. <http://bit.ly/Data4SustDev>
- World Economic Forum (2013) Green Light Creating the Business Case for CITES: A New Finance Mechanism. *Global Agenda Council on Governance for Sustainability*. WEF, 11 p. https://cites.unia.es/cites/file.php/1/files/sustainability_greenlight.pdf (Accessed 22 Feb 2019)
- Zacharias M. (2014) *Marine Policy. An introduction to governance and international law of the oceans*. Earthscan Routledge, London, xxii+313 pp.

Appendix. List of organizations identified as named entities in international environmental law conventions, using SPACY and CoreNLP software (list produced in 2018 by E. Piccinini et

C. Rohrbacher, and after a manual screening of the raw results by P. Mazzega and C. Lajaunie).

1. ASCOBANS Secretariat of the Agreement on the Conservation of Small Cetaceans of the Baltic, North East Atlantic, Irish and North Seas
2. Ad Hoc Study Group on Indicators of Coral Bleaching
3. Agreement on the Conservation of Albatrosses and Petrels
4. Agreement on the Conservation of Cetaceans of the Black Sea
5. Alliance of Small Island States
6. American Tropical Tuna Commission
7. Aquatic Mammals Working Group
8. Arctic Council
9. Arctic Council's Conservation of Arctic Flora and Fauna
10. Arctic Migratory Birds Initiative
11. Atlantic Flyway Shorebird Initiative - CMS
12. Atoll Ecosystem Conservation
13. Australian Institute of Marine Science
14. Barcelona Convention for the Protection of the Marine Environment and the Coastal Region
15. Benguela Current Commission
16. Benguela Current Large Marine Ecosystem Program BCLME
17. Bycatch Working Group - ASCOBANS
18. Cahuita National Park
19. Caribbean Community CARICOM
20. Central Pacific Fisheries Commission
21. Coast Turtles MoU
22. Commission for the Conservation of Antarctic Marine Living Resources CCAMLR
23. Commission for the Conservation of Southern Bluefin Tuna CCSBT
24. Common Wadden Sea Secretariat
25. Convention for the Protection of the Marine Environment of the North-East Atlantic OSPAR Commission
26. Convention on the Conservation of Antarctic Marine Living Resources CCAMLR
27. Convention on the Prevention of Marine Pollution by Dumping of Wastes
28. Coral Triangle Center
29. Council of the International Seabed Authority
30. Division for Ocean Affairs and the Law of the Sea UNDOALOS
31. Dolphin Conservation Society
32. Dorob National Park
33. Ecosystem Management-Fisheries Expert Group - IUCN
34. Fisheries Expert Group - IUCN
35. General Fisheries Commission for the Mediterranean GFCM
36. Gandoca Manzanillo National Wildlife Refuge

37. GloBallast Partnerships Programme
38. Global Coral Reef Database
39. Global Coral Reef Monitoring Network Management Group GCRMN
40. Global Ocean Biodiversity Initiative GOBI
41. Global Ocean Observing System GOOS
42. Governing Council and Regional Seas Programme
43. Great Barrier Reef Marine Park Authority GBRMPA
44. Hainan Dongzhaigang Mangrove National Natural Reserve
45. IUCN Shark Specialist Group
46. Indian Ocean Tuna Commission
47. Integrated Coastal Zone Management Project ICZM - WB
48. Integrating Watershed and Coastal Areas Management in Caribbean Small Island Developing States IWCAM
49. Intergovernmental Oceanographic Commission UNESCO
50. International Center for Living Aquatic Resources Management ICLARM
51. International Commission for the Conservation of Atlantic Tunas ICCAT
52. International Coral Reef Action Network ICRI network
53. International Coral Reef Initiative ICRI
54. International Council for the Exploration of the Sea ICES
55. International Maritime Organization IMO
56. International Plan of Action for Sharks IPOA SHARKS
57. International Seabed Authority ISA
58. International Society for Mangrove Ecosystems
59. International Whaling Commission IWC
60. Korea Maritime Institute
61. Mediterranean Protected Areas Network MEDPAN
62. Marine Environmental Protection Committee - IMO
63. Maritime Innovative Territories International Network
64. Marine Stewardship Council
65. Marine Strategy Framework Directive Task Group – European Commission
66. Marine Trophic Index
67. Mediterranean Protected Areas Network
68. Monaco Blue Initiative
69. Nairobi Convention Coral Reef Task Force
70. Nairobi Convention Secretariat
71. Nairobi Framework Partnership
72. Namibia Islands Marine Protected Area NIMPA
73. National Oceanic and Atmospheric Administration NOAA
74. North East Atlantic Fisheries Commission NEAFC
75. North Pacific Marine Science Organization
76. North-East Asian Marine Protected Areas Network NEAMPAN

77. Northwest Atlantic Fisheries Organization NAFO
78. Ocean Biogeographic Information System OBIS
79. Ocean Declaration
80. Ocean Health Index
81. Ocean Teacher Global Academy
82. OSPAR Commission ocean
83. Pacific Loggerhead Turtle Action Plan
84. Permanent Commission for the South Pacific CPPS
85. RINA Services S.p.A. marine
86. Reef Resilience Network – Nature Conservancy
87. Réseau Régional d'Aires Marines Protégées d'Afrique de l'Ouest RAMPAO
88. Sea Mammal Research Unit
89. Seabird Bycatch Working Group SBWG - ACAP
90. Secretariat of Convention on the Conservation of Antarctic Marine Living Resources
CCAMLR
91. Secretariat of the Pacific Community SPC
92. Secretariat of the Pacific Regional Environment Programme SPERP
93. Shankou Mangrove National Nature Reserve
94. Society for Marine Mammalogy
95. South Pacific Regional Fisheries Management Organisation SPRFMO
96. Sustainable Ocean Initiative - CBD
97. Suvarrow National Park
98. Tioman Marine Park
99. UN Fish Stocks Agreement
100. United Nations Convention on the Law of the Sea UNCLOS
101. United Nations Fish Stocks Agreement
102. United Nations Open-ended Informal Consultative Process on Oceans and the Law of
the Sea UNICPOLOS
103. United States Coral Reef Task Force
104. Western Hemisphere Shorebird Reserve Network WHSRN
105. Western Indian Ocean Coastal Challenge WIOCC
106. Western Indian Ocean Marine Science Association WIOMSA
107. Western and Central Pacific Fisheries Commission WCPFC
108. Wider Caribbean Sea Turtle Conservation Network WIDECAST
109. World Fish Center
110. World Fisheries Trust