Article

A Big Data View of the European Energy Union: Shifting from 'a Floating Signifier' to an Active Driver of Decarbonisation?

Karoliina Isoaho *, Fanni Moilanen and Arho Toikka

Department of Social Research, University of Helsinki, 00140 Helsinki, Finland; E-Mails: karoliina.isoaho@helsinki.fi (K.I.), fanni.moilanen@helsinki.fi (F.M.), arho.toikka@helsinki.fi (A.T.)

* Corresponding author

Abstract

The Energy Union, a major energy sector reform project launched by the European Commission in 2015, has substantial clean energy and climate aims. However, scholarly caution has been raised about their feasibility, especially with regards to accommodating climate objectives with other closely related yet often competing policy goals. We therefore investigate the policy priorities of the Energy Union by performing a topic modelling analysis of over 5,000 policy documents. A big data analysis confirms that decarbonisation and energy efficiency dimensions are major building blocks in the Energy Union's agenda. Furthermore, there are signals of policy convergence in terms of climate security and climate affordability policies. However, our analysis also suggests that the Commission is not actively prescribing trajectories for renewable energy development or paying close attention to declining incumbent energy generation technologies. Overall, we find that the Energy Union is not a 'floating signifier' but rather has a clear and incrementally evolving decarbonisation agenda. Whether it further develops into an active driver of decarbonisation will largely be determined by the implementation phase of the project.

1. Introduction

Decarbonising the energy sector and shifting towards clean energy is high on the agenda in the European Union (EU). A major demonstration of the political will to integrate climate and energy objectives came in February 2015, when the European Commission (hereafter, EC or the Commission) unveiled its blueprint for launching a 'European Energy Union'. While the Energy Union project was set to push the European energy market towards better security, sustainability and affordability, it also included novel substantial elements to transform the EU's energy system into one working towards clean energy (Siddi, 2016, p. 131).

However, despite being applauded for having the potential to fundamentally transform Europe's energy system (Ringel & Knodt, 2018), the Energy Union has also received scholarly criticism. This is because the Energy Union package comprises numerous policy goals and the emphasis given to each dimension still remains an open question (Szulecki, Fischer, Gullberg, & Sartor, 2016). Caution has also been specifically raised in terms of the EU's decarbonisation ambitions. Prior research has highlighted that despite climate gaining traction as a topic at the EU level, actual climate and energy policy integration in EU legislation has remained difficult (Dupont & Oberthür, 2012; Russel, den Uyl, & de Vito, 2018). It is therefore clear that for the Energy Union to succeed in its

decarbonisation objectives, it will have to face and overcome the challenges of delivering policy convergence in a way that other policy areas, such as security and competitiveness, are not at odds with sustainability goals.

Is the Energy Union transforming the EU's energy policy agenda towards more policy convergence and active promotion of climate and decarbonisation objectives? If it is, this is likely to be visible already in the Commission's energy policy goals, language and structure, as the Energy Union project is proceeding towards its implementation phase from 2019. Motivated by this knowledge gap in the literature, and in light of the concern about the Energy Union's policy ambiguity identified in prior research, this article looks at large-scale energy policy document corpora before and after the launch of the Energy Union. We investigate the priorities given to the different policy dimensions and identify potential signs of policy convergence. To do so, we perform a topic modelling analysis on over 5,000 policy documents. While many studies have examined the EC's decarbonisation discourse and agenda at the conceptual level or based on smaller qualitative data sets (Benson & Russel, 2015; Schreurs, 2016; Szarka, 2016), taking a big data approach allows us to gain a novel perspective of the Energy Union. We investigate changes at the policy document level at a scale that has been beyond reach for previous studies that used other methods. As such, the results will offer valuable stepping stones to further research and policy.

The article unfolds as follows. Section 2 briefly introduces the climate and energy policy development in the EU and discusses the agenda shaping power of the Energy Union project. Section 3 outlines the analytical approach used in this article. Section 4 explains the steps taken in our data collection and analysis, while section 5 presents the results. Finally, we discuss our findings in section 6 and give concluding remarks in section 7.

2. An Energy Union in the Making

2.1. Climate and Energy Policy Nexus in the EU

Energy policy in the EU has remained exclusively under the control of national administrations (Benson & Russel, 2015). While energy policy is a shared competence between the EU and its member states, the latter retain the right to decide on their respective energy mixes as per Article 194 of the Treaty on the Functioning of the EU (EU, 2007). The Lisbon Treaty of 2009 was a first step towards introducing energy policy as an independent issue area at the supranational level (Council of the European Union, 2007). The subsequent years introduced groundwork initiatives towards integrating climate and energy policy objectives into European legislation. In the aftermath of the Lisbon Treaty, the Commission also started to form a discourse about a common European energy policy, featuring security of supply, sustainability and competitiveness as its main pillars (Szulecki et al., 2016). Thus, at the discursive level, the EC made attempts to

link previously separate and oftentimes conflicting policy to create an idea of the energy-climate nexus as 'European in scope' where 'the problem is a common European one, and by extension so is the solution' (Maltby, 2013, p. 437).

The change of presidency in the Commission in 2014, coupled with climate and energy security becoming critical matters in the European policy scene, created a 'window of opportunity' for further integrating climate and energy at the policy and institutional level (Bürgin, 2018). The Energy Union project was officially launched in 2015 by the Commission's new president, Jean-Claude Juncker, as part of the 'ten Commission priorities for 2015–2019' (EC, 2015). The initiative for such a Union was first proposed by Donald Tusk, the previous President of the European Council and the Prime Minister of Poland. Tusk saw the Union as a means to secure the exploitation of member states' fossil fuel reserves, focussing largely on ensuring security of supply. He outlined the first strategies of the Energy Union to be a response to concerns about energy security in Europe following the Ukrainian crisis. However, with the Juncker Commission in charge, the Western European states became the decisive actors behind the Energy Union, and the project was geared towards questions of market integration and clean energy ambitions (Siddi, 2016). The Energy Union was designed to contain five 'closely related and mutually reinforcing dimensions': (1) energy security, solidarity and trust; (2) a fully integrated European energy market; (3) energy efficiency contributing to moderation of demand; (4) decarbonising the economy; and (5) research, innovation and competitiveness (EC, 2015). These dimensions show that the EC's overarching objective is transforming the energy system towards being a clean energy system, parallel to creating one common European framework for energy. In 2016, the Commission published its 'Clean Energy for All Europeans' package, where a more detailed agenda was established for putting energy efficiency first, achieving global leadership in renewable energies (RES) and providing a fair deal for consumers (EC, 2016).

2.2. The Agenda Shaping and Institutional Power of the Energy Union Concept

Despite the Commission's efforts to put its decarbonisation plans into practice, the Energy Union project has been received with caution by many scholars. As Szulecki et al. (2016, p. 584) aptly summarise, while the Energy Union is a significant policy idea, it runs the risk of remaining 'a floating signifier' or 'an empty box in which every stakeholder tries to put whatever is on the top of their priority list'. This is partly because the Commission did not specify the order of priority for the five Energy Union dimensions to begin with. In other words, while the primary aim of the Energy Union is to form a streamlined and coherent energy policy framework, its nature and ambition has been left open for interpretation. There is

a risk that, when multiple political groups try to push the project to reflect their priorities and demands, the Energy Union absorbs rather than emits meaning, either becoming 'everything', i.e. signifying an impossible combination of policy ambitions in pursuit of limiting contestation, or 'a floating signifier' (Laclau & Mouffe, 2013), where the concept is non-fixed in its meaning and can thus be used to advance different policy objectives and priorities depending on the context.

Why is the Energy Union's conceptual development of importance? When drawing insights from the EU institutional and governance scholarship, two interlinked points come to the fore. First, by institutionalising the European energy policy through the Energy Union, the Commission is working actively as an energy 'policy entrepreneur' to shape the energy policy agenda on the European continent (Tosun, Biesenbender, & Schulze, 2015). We refer to 'agenda shaping' here in line with Tosun et al. (2015) as a process through which certain images are emphasised or reemphasised to introduce new issues, restructure the existing strategies and plans, or bar issues from the political agenda. Through agenda shaping, the EC is creating the ideational structures for the Energy Union, which in turn play a central role in determining which issues are considered significant and legitimate in the new institutional setting (Carstensen & Schmidt, 2016). The way the EC decides to promote energy policy through the Energy Union concept also has ramifications for institution building. For instance, scholars have noted that as a policy entrepreneur, the EC also steers the implementation rules for policy, thereby affecting institutional structures and their potential effectiveness (e.g. whether the emphasis is on depth or participation) (Bernauer, Kalbhenn, Koubi, & Spilker, 2013).

Second, the conceptual development of the Energy Union project is important because it highlights the challenges to institutionalise a common European agenda when member states hold power over their domestic energy policy. This will not be feasible unless the voice of the major member states and member state blocks are taken into account (Tosun et al., 2015, p. 7). Scholars have already highlighted sticking points in the Energy Union's formation as member states seek to influence the Commission's agenda by shaping activities to further their domestic policy priorities (Bürgin, 2018; Szulecki et al., 2016). These include, for example, a debate surrounding the role of energy efficiency in the Energy Union's agenda. A block of member states has viewed efficiency questions as being at odds with affordability, especially as over 70 per cent of the existing building stock remains highly inefficient and, thus, costly to upgrade (Ringel & Knodt, 2018). As a result of this controversy, energy efficiency was even omitted from the agenda of some of the meetings surrounding the launch of the Energy Union (Siddi, 2016). Another major challenge has to do with promoting decarbonisation objectives. Many Central and Eastern European countries have wanted to take advantage of the Energy Union to im-

prove their own energy security and exploitation of domestic resources (coal and shale gas), whereas the Western member states have had a clear preference for advancing climate policy (Bocquillon & Maltby, 2017).

It is also important to note that the Energy Union's agenda shaping is not done in a vacuum between member states and EU institutions. The agenda of the Energy Union is also susceptible to international influence, like the Paris Climate Agreement, and driven by external events like the Ukrainian crisis or economic recession (Szulecki et al., 2016). In this article however, we are less interested in the causality in agenda shaping—i.e., what is influencing the Energy Union's formation—but rather focus on exploring the substance and shape of the Energy Union's concept and related agenda.

3. A Text-as-Data Approach to the Energy Union

We approach the Energy Union project with a text-as-data approach. These refer to research designs in which texts are analysed statistically with different degrees of automation. For machine coding, both open-ended and a priori defined methodologies are available. In machine learning terminology, these are referred to as unsupervised and supervised methods. While supervision means that the method is supplied with examples to learn from, unsupervised machine-learning methods are solely based on associations between terms in the corpus itself (Toikka & Purhonen, 2016). The choice of a computational method is always connected to the research question, and there is no one globally best method to use—but rather various different tools to augment human interpretation (Grimmer & Stewart, 2013). While the main goals of the Energy Union are explicit and known, the status of its current agenda is less so. Thus, applying unsupervised methods that reveal latent thematic structure in a corpus is considered the most appropriate approach to gaining insights into the Energy Union's conceptual development.

Specifically, in this article we have chosen to apply the method of topic modelling (Blei, 2012). Put briefly, topic modelling is a collective term given to a family of computational algorithms built on Bayesian probabilistic theory. The intuition behind topic modelling is simple: it refers to a generative process in which an algorithm tries to recreate a given corpus assuming 'that each document is made up of a mixture of topics, as well as a mixture of words associated with each topic' (Boussalis & Coan, 2016, p. 92). In other words, if we ignore all syntax and grammar and simply imagine that the documents were generated by an author picking tokens (words or combinations of words) from topics (probability distributions of the tokens), the topic model estimates the word and token distributions that would most likely have generated the observed corpus. The output contains two items: the topic word proportions (usually presented as a set of word lists) and the document topics proportions (van Atteveldt, Welbers, Jacobi, & Vliegenthart,

2014). These are then subjected to human interpretation. Hence, unlike in supervised or qualitative text analysis techniques, topic modelling shifts interpretation of data to a later stage as the computer is allowed to categorise all data before human intervention. It is important to note that it is not given that all the word clusters generated as topic output clearly represent content categories of the policy context analysed. Some topics may represent multiple issues or even non-policy-related areas. Therefore, it is pivotal for the researchers to reflect critically on the results against their empirical case and pay close attention both to the words in a topic and the documents most associated with these words when interpreting the topics.

The particular algorithm we apply is Latent Dirichlet Allocation (LDA; Blei, Ng, & Jordan, 2003), which is the most-used and is a well-established algorithm in text analysis. The main function of LDA is to provide an automated way to explore 'the presence of meaningful clusters of terms' running through a corpus (Boussalis & Coan, 2016, p. 92). As words in these clusters—topics—relate to one another through word co-occurrence and probability distributions, topic modelling analysis is argued to have 'high levels of substantive interpretability' (DiMaggio, Nag, & Blei, 2013) and has been found to be effective in applied settings in social and political science (Nowlin, 2016).

In our case, there are several reasons why we find applying topic modelling to policy document analysis particularly relevant. First, topic modelling gives us the benefit of scale and scope for the analysis as it allows for coverage of more data than researchers could qualitatively read. Second, deconstruction of large policy corpora is also important because it helps reveal how specific policy questions are presented by policy-makers. This is because policy language can be seen as expressing political purpose and intended courses of action by policy-makers (Majone, 1989). Third, taking a topic modelling approach to policy document analysis is insightful in the EU context. As 'a policy entrepreneur', the EC has agenda-shaping power vis-à-vis the Energy Union (Bürgin, 2018). In this light, the policy documents the EC publishes are likely to yield important signals of the ways in which the Commission is constructing its representation regarding a problem and, in our case, regarding decarbonisation.

Indeed, harnessing a topic modelling approach also comes with limitations. It is clear that policy documents alone do not exhaustively cover the policy challenges surrounding decarbonisation, but rather they provide a focalised account of the descriptive phases in the policy cycle (Knill & Tosun, 2008). It is also important to acknowledge that the topic model technique does not intend to fully utilize all the rich information available in political texts but rather to reveal latent structure in a corpus. Nonetheless, while the topic model does not take into account semantics, syntax or the order of words in the documents, topics have been found to correspond to similar substantive issues when compared

against manually-annotated themes (van Atteveldt et al., 2014). In our study, we interpreted topic output as signals of policy ideas and issues. Topic models provide us with an aggregate picture of the latent thematic structure of the EC's policy documents—a snapshot of the policy documents—which is useful in exploring the thematic formation of the Energy Union. For a more in depth account about how to interpret topics for the purposes of political science research, see for example Chang, Gerrish, Wang and Blei (2009) and van Atteveldt et al. (2014).

Finally, the reader should also be informed that there are also extensions of the LDA approach which have been developed to improve the simple LDA by including document metadata into the analysis (e.g., author metadata with author-topic models, time metadata with dynamic topic modelling, generalized to any number of factors as with factorial LDA). These extensions are definitely useful for numerous research questions, especially when relevant metadata related to the issues of interest are available. However, our aim is to offer the reader an empirical exploration of the energy policy documents rather than claim to reveal causal relations or the like. Thus, for both model parsimony (use the simplest model that answers your questions) and the sake of presentation, we decided that a simple LDA would be sufficient for the analysis here.

4. Data Collection and Analysis

In our approach, we compared the thematic structure of two policy document corpora. The first one comprises documents written and published prior to February 2015, i.e., the official launch of the Energy Union, and therefore represents the EC's policy path prior to the Energy Union. The earliest documents in the pre-corpus date to 2001, although 90% of the corpus has been published between 2009 and 2014. The second corpus contains documents developed under the Energy Union agenda, from the period of 2015–2018. As the Energy Union was given a four-year schedule (2015–2019), most of the agenda setting forming the thematic development of the Energy Union will have taken place prior to 2019.

4.1. The Two Corpora

The data set consists of 5,055 documents from the Commission's Directorate-General for Climate Action (DG Clima) and from the Directorate-General for Energy (DG Energy) websites. These sources were chosen as they regroup the main EU-level documents for energy and climate-related matters. The data were downloaded in April 2018 and collected as follows: we started from the energy-relevant main pages and followed all internal links as long as they stayed within the domain, downloading all PDF documents available.

Table 1 presents our data set. We coded document year manually to guarantee correct timestamps and

assigned each document a category using the EC's official list of document types. These were supplemented by our own categories when documents did not fall into any of the official ones. As Table 1 shows, the bulk of our data set consists of EU regulations, member states' reporting documents and research publications. Items facilitating communication, like meeting agendas, communications and such, are also a major category. These include two categories written by outside actors as stakeholder and consultation inputs (SI, CI). While we have not systematically sought to include input from the public into the data set—these are generally published elsewhere on the EU websites—the DGs have occasionally published these documents as supplements to meeting information etc., and so we have not removed them. We treat all the documents in the same way regardless of their type—we do not e.g. give priority to regulation as we simply look at how thematic content is shared among different documents. After several rounds of data collection and categorization rechecks by two researchers, we divided the documents into two corpora.

Table 1. Data set. Table represents document types and allocation for each corpus.

		Documents in pre-Energy Union corpus	Documents in post-Energy Union corpus
COM	Proposed legislation and other Commission communications to the Council and/or the other institutions, and their preparatory papers. Commission documents for the other institutions (legislative proposals, communications, reports, etc.)	**177**	**218**
SWD	Commission staff working document	**127**	**184**
SEC	Documents which cannot be classified in any of the other series	**101**	**2**
C	Documents relating to official instruments for which the Commission has sole responsibility. Some are transmitted to the Council or Parliament for information.	**196**	**26**
Communication paper (CP)	Communication to the public, for example flyers, information sheets, PowerPoint presentations etc.	194	202
Country report (CR)	Documents which discuss a certain member state or which are produced by a member state	884	584
Research paper (RP)	Research and document papers authored by commission officials, JRC, researchers or consultants	331	330
Strategy paper (SP)	EU strategies and future-orientated planning	28	20
Meeting paper (MP)	Documents related to a certain project or meeting in DG Clima or DG Energy	472	199
Policy Implementation (PI)	Practical guidance for the implementation of the Commission's policy	68	30
Law (L)	Commission's law publications from *Official Journal of the European Union*	329	136
Communication (CJOURNAL)	Commission's communication publications from *Official Journal of the European Union*	37	41
Stakeholder input (SI)	Stakeholders' independently expressed opinions on policy	1	85
Consultation Input (CI)	Stakeholders' opinions on policy when requested by a consultant	1	52
Overall no. of documents		**2,946**	**2,109**

Notes: The classification of documents was done in accordance with (a) the Commission's categories, which are highlighted in bold and (b) supplementary categories developed by the authors. The EC categories are available at: https://ec.europa.eu/transparency/regdoc/?fuseaction=helpcote&language=en

4.2. Data Analysis

Topic modelling, like any other natural language-processing method, requires pre-processing of the data: the raw files need to be tokenised or transformed into units of analysis. We have chosen to use single words or unigrams in our analysis. Although it can be more appropriate to generate the bag of words using multiword phrases (essentially, model the probability of 'climate change' instead of 'climate' and 'change' separately), topic modelling can produce better results with the simplest unigram tokens (Yau, Porter, Newman, & Suominen, 2014). We used tidytext (Silge & Robinson, 2016), and topicmodels (Grün & Hornik, 2011) R packages for data handling and modelling. To generate the tokens, we removed too common words using stopword from the Snowball English language stopword list, technical terms related to EU documentation (like the repeating EN on every page signifying that this is the English version), too rare words using a cut-off point for the minimum number of documents to include a word, and removed any punctuation, numbers and non-alphabetical characters. We transformed each text into a row on a document-term matrix, summing how often each word in the corpus is used in the document.

It is the researcher's task to decide in advance how many topics the model should produce as its output. While some measures have been proposed for deciding on the number of topics (Chang et al., 2009), there is no standardised one-fits-all procedure for researchers to use. Thus, we followed a popular qualitative approach, which is to define the number of topics based on semantic validity (Quinn, Monroe, Colaresi, Crespin, & Radev, 2006). To gauge the extent to which each topic represents a distinct theme and is coherent in meaning, we ran topic models varying the number of topics and the cut-off point for rare words. We tried 10, 20 and 30 topics, requiring each word to appear in a minimum of 10, 15, and 20 per cent of documents. To compare the models on semantic validity, three researchers independently gave descriptive labels to the topics generated by the test runs by looking at the 20 most probable words and by consulting the 10 most probable documents associated with each topic. Based on the researchers' discussions, we saw that 30 topics gave a good balance: it enabled the model to distinguish clear themes (no themes seemingly discussing two issues) yet did not allow for topics that were too specific to emerge based solely on words associated with a writing style or genre, for example.

The topic interpretation, in other words, assigning a label to each word list, was done by two researchers based on looking at the 20 most probable keywords and consulting the 10 most probable documents associated with the given topic. In Appendix 1 we illustrate the topic interpretation process with examples and in Appendix 2 we present the descriptive name or 'topic label' for all the topics in both corpora, along with the 10 most probable words from each topic. Here we want to highlight that the labelling of topics was done with no prior categorisation in mind. After the interpretation process, we saw that the large majority of topics fell under the broad categories of the Energy Union dimensions (the exceptions being topics about terminology and regulation, see Table 2). We therefore decided to use the five dimensions as an analytical and presentational aid. Using the Energy Union's dimensions also allowed us to compare our results against the official goals of the project, which worked as a form of additional step to guarantee the semantic coherence of our results. As to how the topics were categorised under these dimensions, one researcher compared each topic label (words and top 10 document content) with how the Commission's Communication document (EC, 2015) had defined the dimensions for the Energy Union.

5. Results

To compare topic structure pre- and post-corpus, we grouped the topics in accordance with the five Energy Union dimensions outlined by the EC (see Section 2.1). The topics are presented in Table 2, and their thematic emphasis is illustrated in Figure 1.

By looking at Figure 1, it is clear that the energy efficiency and decarbonisation topics are the most predominant policy issues, covering 50 per cent of the policy themes in the pre and 63 per cent in the post-corpus. The latter corpus also seems slightly more thematically oriented as it has fewer general themes based on regulation and terminology than the pre-corpus. To evaluate changes in the policy documents language under the Energy Union project, we next open up and compare the semantic structure of key decarbonisation and climate-related topics in more detail.

Under the energy efficiency dimension, both corpora exhibit multiple topics on national reporting in the context of the respective time periods' directives. The most probable documents commonly relate to the National Energy Efficiency Action Plans (NEEAs) under the Energy Efficiency Directive 2012/27/EU. Similarities can also be seen with topics that deal with setting standards for transport sectors. However, the pre-corpus topics are about emissions reduction in transport (pre12, pre13), while the post-corpus has a clear emphasis on engine, fuel and combustion efficiency (post15, post26). There are three main differences between the corpora: cogeneration in combined heating and power (CHP) emerged as a solid, new theme (post10, post13); building efficiency is exclusively about the renovation of existing buildings in the post-corpus; and consumers are more present in the post topics, whether through eco-labelling schemes (post9) or considerations over energy costs and poverty among the EU households (post18).

With regards to the decarbonisation dimension, international climate agreements by the United Nation's Framework Convention on Climate Change (UNFCCC) feature in both corpora (pre14, post17), accompanied by
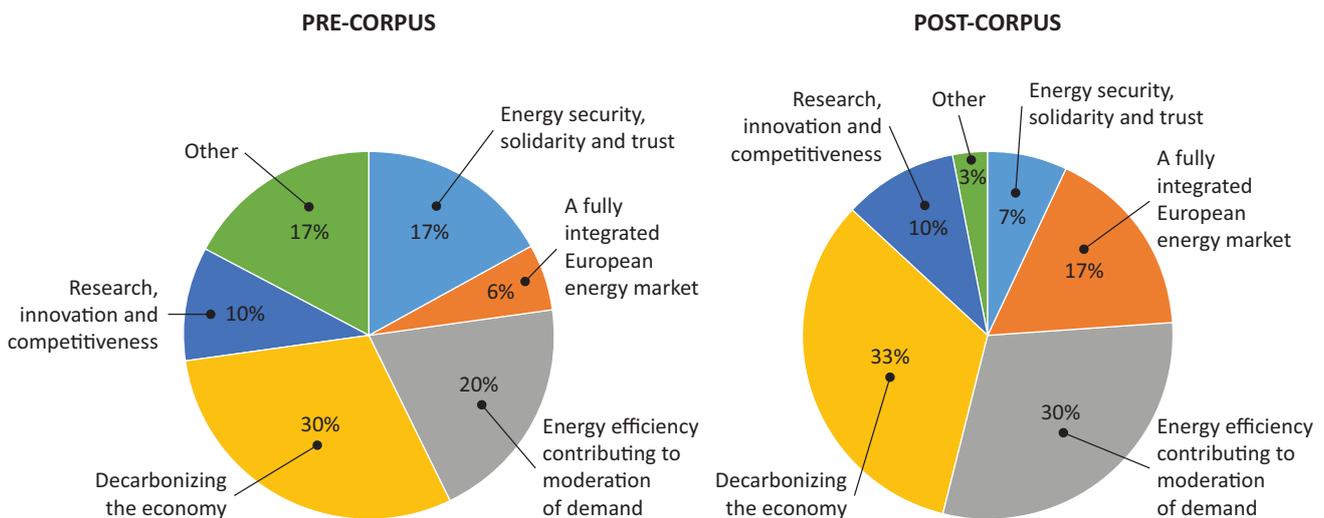
PRE-CORPUS **POST-CORPUS**

Figure 1. Comparison of pre and post-corpus topic structure.

topics on related reporting on member states' annual GHG inventories (pre15, pre17, pre18, post18). However, the model clearly estimates a shift in emphasis to the regulation on binding annual emissions reductions by member states, which come into effect through the Effort-Sharing Regulation and include sectors, such as transport or buildings, which are not covered by the emissions trading scheme (ETS) (post19). Furthermore, 'renewable energy use and promotion' emerges as a similar topic in both corpora, mostly covering progress reports on the National Renewable Energy Action Plan (NREAs). The post-corpus also has one additional topic for biomass and biofuel use and production (post21). The EU's strengthened interest in the ETS policy instrument is also visible (pre19, post 22). Moreover, the pre-corpus has a topic for external climate support, like providing technology transfer and adaptation finance for developing countries (pre22), but these have been replaced in the post-corpus by internal security and risk considerations on guaranteeing security of supply through diversification (post20, post26).

As to the other three dimensions, energy security topics feature slightly less in the post-corpus, whilst market integration has more emphasis compared to the pre-corpus. Emphasis on research and innovation is similarly marginal in both corpora. For energy security, we see topics that pertain to establishing the Euratom Community framework and radioactivity monitoring and nuclear waste management in both corpora (pre1, post1). The model also shows a clear topic on natural gas before and after the Energy Union. The pre-corpus topic is related to documents on guaranteeing security of gas transmission with the EU's external partners, while the post-corpus topic has an emphasis on the role of gas in the internal energy market, with key words on prices and capacity (pre2, post2). Furthermore, the post-corpus is strikingly more focalised on the mandate of delivering an internal market than the pre-corpus. This is visible from topics dealing specifically with the ways the power system can be

organized under a smarter system based and consumer-led Energy Union (post3, post7) and how market failures in such a system can be corrected through instruments such as cross-border capacity mechanisms (post6). Finally, both corpora display themes on research on moving to a competitive low-carbon economy. The post corpus exhibits the EU's 2016 reference scenarios, one of the Commission's key tools for climate analysis. Even if topics in the pre-corpus deal with the abatement costs in the transportation sector, the post27, post28 and post29 topics do even more so, with macroeconomic and technical analyses on decarbonisation scenarios.

6. Discussion

An LDA examination of policy topics does not reveal radical changes in the disposition of the Commission's energy policy priorities after the launch of the Energy Union. Rather, it appears that the Commission has reiterated and strengthened the focus on the energy efficiency and decarbonisation dimensions. Nonetheless, the topic structure does point to interesting incremental changes within climate and decarbonisation policy in the years leading up to the implementation phase of the Energy Union.

First, a critical discussion that our results allude to has to do with the Commission's ambition over RES and the related decarbonisation measures. While the commitment to decarbonise is clear from many topics, especially from the emphasis to assign binding emission reduction regulations on the member states, our findings point to a large degree of 'technology-neutrality' in the Commission's decarbonisation strategy (Fitch-Roy, Benson, & Mitchell, 2018). That biomass is the only standalone RES topic and renewables are otherwise tackled under an umbrella term 'renewable energy promotion and use' suggests that the Energy Union has a less prescriptive approach to RES policy development than the initial Energy Union communication leads one to believe. Promoting

Table 2. Topic labels in pre and post-corpus, categorised by Energy Union dimensions.

Energy Union dimensions	Pre-Corpus	Post-Corpus
Energy security, solidarity and trust	• Nuclear safety and radioactivity management (pre1) • Natural gas issues (pre2) • Electricity generation (pre3) • Technical security (pre4) • Operational regulations (pre5)	• Nuclear safety and radioactivity management (post1) • Gas and security of supply (post2)
A fully integrated European energy market	• Internal energy market (pre6) • Smart grid development (pre7) • Financing Projects of Common Interest (post5) • EU energy regulation networks and cooperation (post6) • Energy consumers and smart grid (post7)	• Internal energy market (post3) • Mainstreaming climate policy into EU structural funds (post4)
Energy efficiency contributing to moderation of demand	• Member state energy efficiency plans (pre8) • Energy efficiency strategy (pre9) • Energy efficiency directive in public buildings (pre10) • Nearly zero emissions buildings (pre11) • Vehicle emissions policy (pre12) • Emission reduction implementation in vehicles (pre13)	• Member state energy efficiency achievements (post8) • Member state strategy for energy renovations in buildings (post9) • Energy efficiency infrastructure (post10) • Energy efficiency and CHP (post11) • CHP and efficient heating and cooling systems (post12) • Energy labelling and eco-design for products (post13) • Household energy prices and costs (post14) • Emissions standards for light vehicles (post15) • Emissions MRV & ARV from maritime transport (post16)
Decarbonizing the economy	• Achievements of Kyoto objectives (pre14) • Member states' GHG inventories (pre15) • Renewable energy promotion and use (pre16) • Member state climate policy assessment (pre17) • Member state support and guidance for emissions reduction (pre18) • Improving the ETS (pre19) • Sustainable land-use (pre20) • Aviation regulation (pre21) • External climate support (pre22)	• Implementing Paris Agreement's GHG targets (post17) • Member state's GHG inventories (post18) • Consultation on Effort Sharing Decision (post19) • Renewable energy promotion and use (post20) • Biomass and biofuel use and production (post21) • Revision of the ETS (post22) • Auctioning of emissions allowances (post23) • Sustainable land-use (post24) • Aviation compliance to GHG reduction and Projects of Common Interest (post25) • Transport fuel quality (post26)
Research, innovation and competitiveness	• Future emissions reduction scenarios (pre23) • Roadmap vocabulary (pre24) • Modelling methodologies (pre25)	• Economic analysis of energy policies and technologies (post27) • Scenarios for transport energy use and emissions (post28) • Competitiveness of transport sector (post29)
Other	• Monitoring and verification (pre26) • Project language (pre27) • Energy measurement vocabulary (pre28) • Energy regulation and directives (pre29) • Common energy terminology (pre30)	• Implementing regulation on energy issues (post30)

Notes: See Appendix 2 for most associated words with labels.

a more technology-neutral approach is a way of granting member states different capacities and resources flexibility in how to approach their clean energy development at the national level. Nevertheless, at the same time, literature on clean energy transitions clearly argues that transforming different sectors towards sustainability requires deliberate steering from public policy, especially with technology-specific targets (Markard, 2018). Moreover, techno-neutrality runs a certain risk of leaving loopholes for RES policy development. Leaving room for interpretation on how to approach RES may encourage the development of less-polluting fossil fuels and unproven or contested technologies. For example, our findings indicate that RES use is largely discussed through biomass and biofuel production in the Energy Union. While it is a renewable energy, research has found that a widespread reliance on biomass and biofuels by the member states is not unproblematic in efforts to mitigate climate change (Soimakallio, Saikku, Valsta, & Pingoud, 2016).

However, it is interesting to note, parallel to techno-neutrality, that our findings also contain many topics on reporting mechanisms (the NEEAs and NREAPs). Indeed, the emergence of these standard and often mandatory reporting documents is not surprising, yet, they could suggest that the EC is prioritising soft governance mechanisms aimed at bringing RES and energy efficiency development goals up to their potential. Namely, it has been noted that the EC is actively using the reporting documents as policy tools to structure dialogue between the Commission and the member states (Ringel & Knodt, 2018). In this way, then, the topic structure suggests that even with a techno-neutrality approach, the EC is exercising soft power through measures of dialogue building and norm creation to promote RES development.

Second, our findings can also be reflected against discussions on potential policy convergence in the Energy Union (Szulecki et al., 2016). Interestingly, we find that energy security issues remain prevalent in the topic structure, even if there are fewer independent topics for energy security in the post-corpus. The topic model estimates that reducing dependence on oil and wider energy supply security are being increasingly dealt with under the energy efficiency and decarbonisation dimensions. For example, the focus on an energy-efficient transport sector in the Energy Union is communicated in terms of helping break oil dependence in the EU. Furthermore, our findings show that the EC's declarations about 'putting energy efficiency first' in its decarbonisation agenda has, to a large extent, trickled down to its policy document language. Although previous research has highlighted the efficiency-affordability debate as one of the major sticking points in the early discussions on the Energy Union, our results suggest that energy efficiency is not only a major topic in the documents but there are also new topics specifically on the renovation aspect of building efficiency improvement, which has been an area of heated debate. The energy efficiency paradigm also appears to have extended into addressing the affordability aspect of the concept (Cherp & Jewell, 2014) as there is a new element on providing low energy prices for consumers and households and considerations over energy poverty in the post corpus. Overall, these findings suggest that the preferences of the Western member state block on climate measures have been prioritized by the Commission to achieve policy convergence and gives ground to argue that the Energy Union is having a streamlining effect on climate-security as well as efficiency-affordability policy.

Third, we find that external climate support was replaced by internal security and security of supply considerations in the post-corpus. As foreign policy instruments on climate finance and adaptation have largely disappeared, at first glance, it appears that the ambitions to 'revitalise European energy and climate diplomacy' stated in the Energy Union's strategy have not been advanced in the policy language (EC, 2015, p. 6). While this could imply granting less policy priority to external energy and climate policy in the first phases of the Energy Union, it could also be a result of a process in which the external agenda is reframed in terms of enhancing energy security and competitiveness in the global energy markets. Schulze (2015) has found that external climate policy making is more successful and more widely legitimated by member states if multiple frames are used to promote it.

Finally, it is important to note that, while the model shows a distinct priority for decarbonisation measures, it did not generate any general topics for fossil fuels or specifically for coal. Apart from the EU ETS, which is a demand side-policy aiming to put a cap on carbon emissions and foster clean energy technologies, there is no thematic interest in the regime destabilisation side of energy transitions (Kivimaa & Kern, 2015). The model did not identify any themes on unlocking the carbon-intensive system with the so-called supply-side control policies on incumbent energy generation technology (like regulation on restricting coal or gas use) or considerations on the effects shifting away from fossil-based systems may have on incumbent industries or regional development in member states. It seems, then, that at policy document level, the Energy Union is not inclined towards actively disrupting the existing fossil fuel-based energy systems. This is critical, as a successful transitioning to clean energy is likely to require an orderly exit from incumbent technologies, especially coal (Markard, 2018). In addition, the topic model suggests that support for innovation in energy transition does not emerge as a priority: the topics falling under the 'research, innovation and competitiveness' dimension are few and mostly deal with transportation issues, with limited focus on how to achieve the outlined pathway to delivering significant emissions reductions by 2030. In summary, while the topic model analysis gives cautious support for viewing the Energy Union as an active driver of decarbonisation, our results point to some degree of scepticism regarding the transformational effect of this agenda.

7. Conclusions

This article has examined the European Energy Union's decarbonisation priorities by taking a big data approach to policy document analysis. The early stages of the Energy Union have sparked criticism over the project being an 'empty box' (covering everything and thus losing meaning), or a 'floating signifier' (where the project would be used to advance conflicting aims). However, we find that in policy language, the decarbonisation and energy efficiency dimensions are clearly major building blocks of the Energy Union, significantly more so than the other three dimensions outlined by the Commission. Hence, the Energy Union does not appear to be empty in meaning, but rather has an increasingly clear decarbonisation agenda.

Furthermore, our results further conceptual knowledge on the formation of the Energy Union. We have shown how the Commission is actively using its agenda-shaping power to promote decarbonisation objectives but that in doing so, it is opting for shallower policy prescriptions as opposed to assigning specific technological and policy solutions for RES. While this can be a sign of the EC wishing to guarantee participation of member states with conflicting agendas, it should also be viewed with caution given the risk of simultaneously encouraging the development of less-polluting fossil fuels. In addition, the Energy Union project has generated further policy convergence between climate, security and affordability. Interestingly, this has been done mainly by restructuring the European energy policy agenda to align it more with the preferences of the Western member states. Finally, and importantly, our results also suggest that the Commission is not paying close attention to phasing out incumbent fossil fuel generation technologies such as coal. Therefore, there appears to be a lack of policy interest in removing the existing high-carbon infrastructures, which is also a critical for a successful decarbonisation of the EU's energy system. As a result, caution remains over the transformational effects of the Energy Union's decarbonisation agenda.

Indeed, whether the Energy Union develops from a 'floating signifier' used to promote competing policy priorities into an active driver of decarbonisation will depend on the subsequent phases of the policy cycle, including successful implementation of the project and active climate action by member states, which our study and the topic modelling method cannot account for. Therefore, future research with more detailed qualitative approaches is needed, among others, on the decarbonisation agenda the Commissions is promoting through its soft-governance mechanisms and on how the aim of solidarity among member states is harnessed for the purpose of advancing climate aims. Tracing the (non-)development of the Energy Union's external climate and energy policy also appears to be a pivotal area for future research, especially given the current context from which we witness a simultaneous delocalisation of European industrial production to developing countries, to areas where pollution is bound to grow and where mechanisms such as adaptation finance and technology transfer would be increasingly needed. Overall, continuing to monitor the EC's policy language will be of particular importance, given the European elections in 2019 and the subsequent change of the President of the Commission, which will put the continuity of the Energy Union's decarbonisation policy under further test.

Conflict of Interests

The authors declare no conflict of interests.

References

Benson, D., & Russel, D. (2015). Patterns of EU energy policy outputs: Incrementalism or punctuated equilibrium? *West European Politics*, *2382*(January), 37–41. https://doi.org/10.1080/01402382.2014.936707

Bernauer, T., Kalbhenn, A., Koubi, V., & Spilker, G. (2013). Is there a "depth versus participation" dilemma in international cooperation? *Review of International Organizations*. https://doi.org/10.1007/s11558-013-9165-1

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77. https://doi.org/10.1145/2133806.2133826

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022. https://doi.org/10.1162/jmlr.2003.3.4-5.993

Bocquillon, P., & Maltby, T. (2017). The more the merrier? Assessing the impact of enlargement on EU performance in energy and climate change policies. *East European Politics*, *33*(1), 88–105. https://doi.org/10.1080/21599165.2017.1279605

Boussalis, C., & Coan, T. G. (2016). Text-mining the signals of climate change doubt. *Global Environmen-*

tal Change, *36*, 89–100. https://doi.org/10.1016/j.gloenvcha.2015.12.001

Bürgin, A. (2018). The impact of Juncker's reorganization of the European Commission on the internal policy-making process: Evidence from the Energy Union project. *Public Administration*, 1–14. https://doi.org/10.1111/padm.12388

Carstensen, M. B., & Schmidt, V. A. (2016). Power through, over and in ideas: conceptualizing ideational power in discursive institutionalism. *Journal of European Public Policy*, *23*(3), 318–337. https://doi.org/10.1080/13501763.2015.1115534

Chang, J., Gerrish, S., Wang, C., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. *Advances in Neural Information Processing Systems*, *22*, 288–296. https://doi.org/10.1.1.100.1089

Cherp, A., & Jewell, J. (2014). The concept of energy security: Beyond the four As. *Energy Policy*, *75*, 415–421. https://doi.org/10.1016/j.enpol.2014.09.005

Council of the European Union. (2017). *Treaty of Lisbon, European Parliament and Council §*. https://doi.org/10.1007/s11552-006-9002-6

DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. *Poetics*, *41*(6), 570–606. https://doi.org/10.1016/j.poetic.2013.08.004

Dupont, C., & Oberthür, S. (2012). Insufficient climate policy integration in EU energy policy: The importance of the long-term perspective. *Earth Systems Governance*, *8*(2). https://doi.org/10.1080/14782804.2014.936370

European Commission. (2015). *A framework strategy for a resilient energy union with a forward-looking climate change policy*. Brussels: European Commission.

European Commission. (2016). *Clean energy for all Europeans*. Brussels: European Commission.

European Union. (2007). Treaty on the functioning of the European Union. *Official Journal of the European Union*. https://doi.org/10.1093/law/9780199684601.003.0029

Fitch-Roy, O., Benson, D., & Mitchell, C. (2018). Wipeout? Entrepreneurship, policy interaction and the EU's 2030 renewable energy target. *Journal of European Integration*, 1–17. https://doi.org/10.1080/07036337.2018.1487961

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, *21*(3), 267–297. https://doi.org/10.1093/pan/mps028

Grün, B., & Hornik, K. (2011). Topicmodels: An R package for fitting topic models. *Journal of Statistical Software*. https://doi.org/10.18637/jss.v040.i13

Kivimaa, P., & Kern, F. (2015). Creative destruction or mere niche creation? Innovation policy mixes for sustainability transitions. *Research Policy*, *2*(1), 29. https://doi.org/10.1016/j.respol.2015.09.008

Knill, C., & Tosun, J. (2008). Policy making. In D. Caramani (Ed.), *Comparative politics* (pp. 495–519). Oxford: Oxford University Press.

Laclau, E., & Mouffe, C. (2013). Hegemony and socialist strategy towards a radical democratic politics. *Journal of Chemical Information and Modeling*. https://doi.org/10.1017/CBO9781107415324.004

Majone, G. (1989). *Evidence, argument and persuasion in the policy process*. New Haven, CT: Yale University Press.

Maltby, T. (2013). European Union energy policy integration: A case of European Commission policy entrepreneurship and increasing supranationalism. *Energy Policy*, *55*, 435–444. https://doi.org/10.1016/j.enpol.2012.12.031

Markard, J. (2018). The next phase of the energy transition and its implications for research and policy. *Nature Energy*. https://doi.org/10.1038/s41560-018-0171-7

Nowlin, M. C. (2016). Modeling issue definitions using quantitative text analysis. *Policy Studies Journal*, *44*(3), 309–331. https://doi.org/10.1111/psj.12110

Quinn, K. M., Monroe, B. L., Colaresi, M., Crespin, M. H., & Radev, D. R. (2006). An automated method of topic-coding legislative speech over time with application to the 105th–108th U.S. Senate. *New York*, 61. Retrieved from http://www.allacademic.com/meta/p150986_index.html

Ringel, M., & Knodt, M. (2018). The governance of the European Energy Union: Efficiency, effectiveness and acceptance of the Winter Package 2016. *Energy Policy*, *112*, 209–220. https://doi.org/10.1016/j.enpol.2017.09.047

Russel, D. J., den Uyl, R. M., & de Vito, L. (2018). Understanding policy integration in the EU—Insights from a multi-level lens on climate adaptation and the EU's coastal and marine policy. *Environmental Science and Policy*, *82*(February), 44–51. https://doi.org/10.1016/j.envsci.2017.12.009

Schreurs, M. A. (2016). The Paris Climate Agreement and the three largest emitters: China, the United States, and the European Union. *Politics and Governance*, *4*(3), 219. https://doi.org/10.17645/pag.v4i3.666

Schulze, K. (2015). Exporting the energy acquis: The external agenda shaping power of the EU. In J. Tosun, S. Biesenbender, & K. Schulze (Eds.), *Energy policy making in the eu building the agenda* (Vol. 28). https://doi.org/10.1007/978-1-4471-6645-0

Siddi, M. (2016). The EU's Energy Union: A sustainable path to energy security? *International Spectator*, *51*(1), 131–144. https://doi.org/10.1080/03932729.2016.1090827

Silge, J., & Robinson, D. (2016). Tidytext: Text mining and analysis using tidy data principles in R. *The Journal of Open Source Software*. https://doi.org/10.21105/joss.00037

Soimakallio, S., Saikku, L., Valsta, L., & Pingoud, K. (2016). Climate change mitigation challenge for

wood utilization—The case of Finland. *Environmental Science and Technology*, *50*(10), 5127–5134. https://doi.org/10.1021/acs.est.6b00122

Szarka, J. (2016). Towards an evolutionary or a transformational energy transition? Transition concepts and roadmaps in European Union policy discourse. *Innovation: The European Journal of Social Science Research*, *29*(3), 222–242. https://doi.org/10.1080/13511610.2016.1166039

Szulecki, K., Fischer, S., Gullberg, A. T., & Sartor, O. (2016). Shaping the 'Energy Union': Between national positions and governance innovation in EU energy and climate policy. *Climate Policy*, *16*(5), 548–567. https://doi.org/10.1080/14693062.2015.1135100

Toikka, A., & Purhonen, S. (2016). 'Big datan' haaste ja uudet laskennalliset tekstiaineistojen analyysimenetelmät: esimerkkitapauksena aihemallianalyysi tasavallan presidenttien uudenvuoden-

puheista 1935–2015 [The challenge of 'big data' and novel statistical methods for textual analysis: A topic model analysis of Presidential New Year Speeches in Finland 1935–2015]. *Sosiologia*, *53*(1), 6–27.

Tosun, J., Biesenbender, S., & Schulze, K. (2015). *Energy policy making in the EU Building the Agenda. Lecture notes in energy*. https://doi.org/10.1007/978-1-4471-6645-0

van Atteveldt, W., Welbers, K., Jacobi, C., & Vliegenthart, R. (2014). LDA models topics....But what are "topics"? *Big Data in the Social Sciences Workshop*. Retrieved from http://vanatteveldt.com/wp-content/uploads/2014_vanatteveldt_glasgowbigdata_topics.pdf

Yau, C. K., Porter, A., Newman, N., & Suominen, A. (2014). Clustering scientific documents with topic modeling. *Scientometrics*. https://doi.org/10.1007/s11192-014-1321-8

About the Authors

Karoliina Isoaho is a PhD candidate in environmental policy at the University of Helsinki, studying the politics of sustainable energy transitions. In her PhD research, Karoliina examines the use of existing (discursive) and emerging (machine learning) textual methodologies to study energy policy in the empirical context of decarbonisation in Europe. Her prior research interests include climate adaptation in the private sector and renewable energy policy in China.

Fanni Moilanen is a graduate student in social and public policy at the Faculty of Social Sciences at the University of Helsinki. Fanni's research interests lie in sustainability, particularly in energy and climate policies at both national and global levels. She enjoys investigating the influence of societal rules and norms, and their change in the society.

Arho Toikka is a university Lecturer in Social and Public Policy at the Faculty of Social Sciences at the University of Helsinki. He is a specialist in environmental policy research, especially energy policy, and the use of novel methodologies in social science, including social network analysis, machine learning, statistical modelling of text data, combinations of qualitative and quantitative approaches.

Appendix 1

1. Topic interpretation

It is critical to consider the reliability and validity of topic modelling output in the interpretation process to guarantee the soundness of results (Grimmer & Stewart, 2013). While computational approaches are reliable in the sense that they will produce the same results each time, validity needs further consideration. As explained in section 4.3 in the manuscript, two authors interpreted the topics independently based on the 20 most probable terms of a topic as well as consulting the 10 most probable documents associated with each topic. Interpreting topic output with these two steps is an intrinsic part of the interpretation process. In addition, consulting the documents words as a validation step for semantic coherence. As two authors independently looked into both of these sources and gave a tentative label to each topic, the authors were able to compare and double-check their findings for semantic coherence. Cases of ambiguity (mostly having to do with the topic exhibiting two issue areas) were discussed among the researchers. Finally, they decided upon the final label to be given for each topic. In Table 1, we provide further evidence on the semantic validity of our results by illustrating the interpretation process. The examples show how we labelled i) topics that only exist in one of the corpora, ii) similar topics and iii) cases of ambiguity.

Table 1. Examples of topic interpretation procedure.

Example	Top 10 words	Extracts from most related documents	Topic label as assigned by authors
Topic in pre-corpus	climate, change, development, policy, action, support, carbon, research, management, sustainable	"The Global Climate Change Alliance (GCCA)…strives to support those poorer countries and regions most vulnerable to climate change by building the human, technical and financial capacity needed to mount—and surmount—the challenge." (*Integrated Climate Strategies to Climate Finance Effectiveness*) "Coordinated by the European Commission, the GCCA provides technical and financial support to developing countries to help them integrate efforts to tackle climate change into their policies and budgets." (*Adaptation Factsheet 2014*)	External climate support
Topic in post-corpus	option, options, costs, eu, cost, european, level, current, annual, national	"The European Commission today launches a public consultation on the preparation of a legislative proposal on the effort of Member States to reduce their greenhouse gas emissions to meet the European Union's emission reduction commitment in a 2030 perspective. It concerns the continuation in the period 2021-2030 of the current Decision 406/2009/EC on the effort of Member States to reduce their greenhouse gas emissions to meet the Community's greenhouse gas emission reduction commitments up to 2020." (*Member States' Consultation Documents on ESD*)	Consultation on Effort Sharing Decision (ESD)
Same topic label in pre and post-corpus	*Pre-corpus:* nuclear, safety, data, protection, waste, risk, management, monitoring, activities, technical	*Pre-corpus:* "It is important to consider the challenge posed by radioactive waste since its management is, and will remain, a long-term issue. Even over the timescale for which surveillance of short-lived wastes is required national borders might change. This, together with potential cross-border impacts means that an international context for radioactive waste management becomes increasingly relevant with the passage of time." (*Sixth Situation Report on Radioactive Waste and Spent Fuel management In the European Union*)	*Pre-corpus:* Nuclear safety and radioactivity management
	Post-corpus: nuclear, safety, waste, management, storage, site, fuel, national, power, protection	*Post-corpus:* "Article 35 of the Euratom Treaty requires that each Member State shall establish the facilities necessary to carry out continuous monitoring of the levels of radioactivity in air, water and soil and to ensure compliance with the Basic Safety Standards1. Article 35 also gives the European Commission (EC) the right of access to such facilities in order that it may verify their operation and efficiency." (*Art. 35 Technical Report—LU 15-04*)	*Post-corpus: Nuclear safety and radioactivity management*

Table 1. (Cont.) Examples of topic interpretation procedure.

Example	Top 10 words	Extracts from most related documents	Topic label as assigned by authors
Same theme but different topic label in pre and post-corpus	*Pre-corpus:* energy, efficiency, electricity, consumption, sector, demand, european, industry, primary, generation	*Pre-corpus:* "Pursuant to Article 24(1) of Directive 2012/27/EU of the European Parliament and of the Council of 25 October 2012 on energy efficiency amending Directives 2009/125/EC and 2010/30/EU and repealing Directives 2004/8/EC and 2006/32/EC, each Member State must by 30 April of each year starting in 2013 report on the progress achieved towards national energy efficiency targets." (*Sweden's annual report pursuant to Article 24 (1) of Directive 2012/27/EU*)	*Pre-corpus:* Energy efficiency strategy
	Post-corpus: buildings, energy, building, renovation, performance, residential, construction, requirements, public, measures	*Post-corpus:* "The existing building stock is the sector providing the greatest potential for achieving energy savings, as buildings account for just over one third of all energy consumed. The Energy Efficiency Directive therefore lays down that Member States should establish a long-term strategy for mobilising investments in the renovation of the national building stock in order to increase the rate of building renovation." (*Slovenia's Long-Term Strategy for Mobilising Investments in the Energy Renovation of Buildings, 2015*)	*Post-corpus:* Member State strategy for energy renovations in buildings
Case of ambiguity whereby the topic contains two themes	united, aviation, france, kingdom, germany, air, operator, spain, european, italy	"Amending Regulation (EC) No 748/2009 on the list of aircraft operators which performed an aviation activity listed in Annex I to Directive 2003/87/EC on or after 1 January 2006 specifying the administering Member State for each aircraft operator." (*Commission Regulation (EU) 2018/336*) "The list of the projects of common interest (PCIs) by country – the (third) Union list of PCIs." (*Regulation (EU) No 347/2013*)	Aviation compliance to GHG reduction and Projects of Common Interest (PCIs)

Appendix 2

Table 2. Topic labels with 10 most probable words. Order of topics insignificant.

	Pre: Topic label with top 10 keywords		Post: Topic label with top 10 keywords
1	**Nuclear safety and radioactivity management** nuclear, safety, waste, management, storage, site, fuel, national, power, protection	1	**Nuclear safety and radioactivity monitoring** nuclear, safety, data, protection, waste, risk, management, monitoring, activities, technical
2	**Natural gas issues** gas, eu, european, commission, market, supply, infrastructure, security, europe, countries	2	**Gas and security of supply** gas, price, market, prices, storage, supply, eu, oil, capacity, natural
3	**Internal energy market** price, market, electricity, prices, total, gas, demand, generation, chp, power	3	**Internal energy market** capacity, market, generation, demand, price, power, cross, markets, electricity, border
4	**Operational regulations** law, oil, environment, act, gas, activities, economic, financial, authority, section	4	**EU energy regulation networks and cooperation** article, commission, directive, regulation, european, union, ec, energy, eu, council
5	**Smart grid development** smart, system, grid, network, data, market, information, management, systems, transmission	5	**Energy consumers and smart grid** electricity, energy, smart, consumers, system, distribution, network, grid, demand, market
6	**Member state energy efficiency plans** allocation, commission, directive, allowances, plan, national, installations, emissions, decision, total	6	**Member state energy efficiency achievements** energy efficiency, savings, measures, consumption, public, article, implementation, saving, measure
7	**Energy efficiency strategy** energy, efficiency, electricity, consumption, sector, demand, european, industry, primary, generation	7	**Member states strategy for energy renovations of buildings** heat, heating, cooling, potential, demand, electricity, district, chp, gas, efficiency
8	**Energy efficiency directive in public buildings** energy, savings, efficiency, measures, consumption, measure, public, sector, saving, buildings	8	**Energy efficiency infrastructure** energy, heating, system, consumption, buildings, heat, electricity, systems, primary
9	**Nearly zero emissions buildings** energy, buildings, building, performance, requirements, residential, heating, renewable, national, article	9	**Energy Efficiency and CHP** energy, heat, consumption, gross, electricity, plants, chp, final, data, generation
10	**Technical security** information, requirements, assessment, report, national, protection, safety, design, risk, response	10	**CHP and efficient heating and cooling systems** buildings, energy, building, renovation, performance, residential, construction, requirements, public, measures
11	**Vehicle emissions policy** emissions, vehicles, km, cars, average, vehicle, target, reduction, diesel, manufacturers	11	**Transport fuel quality** fuel, fuels, diesel, directive, oil, quality, report, biofuels, transport, content
12	**Emission reduction implementation in vehicles** fuel, vehicle, vehicles, emissions, transport, consumption, test, type, cars, road	12	**Emissions standards for light vehicles** vehicle, vehicles, emissions, cars, mass, fuel, reduction, data, average, km
13	**Aviation regulation** allowances, aviation, account, period, trading, operators, operator, volume, emissions, national	13	**Aviation compliance to GHG reduction and Projects of Common Interest** united, aviation, france, kingdom, germany, air, operator, spain, european, italy
14	**Achievements of Kyoto objectives** emissions, eu, climate, change, countries, measures, european, kyoto, greenhouse, emission	14	**Implementing Paris Agreement's GHG targets** eu, energy, emissions, policy, targets, policies, climate, sector, ghg, emission
15	**Member states' GHG inventories** emissions, gas, ghg, emission, report, greenhouse, data, inventory, reporting, carbon	15	**Member state's GHG inventories** reporting, emissions, data, ms, report, review, reported, eu, monitoring, national

Table 2. (Cont.) Topic labels with 10 most probable words. Order of topics insignificant.

	Pre: Topic label with top 10 keywords		Post: Topic label with top 10 keywords
16	**Member state support and guidance for emissions reduction** production, installations, heat, installation, united, allocation, product, emissions, operator, ets	16	**Consultation on Effort Sharing Decision** option, options, costs, eu, cost, european, level, current, annual, national
17	**Renewable energy promotion and use** energy, renewable, res, support, national, sources, electricity, scheme, plan, development	17	**Renewable energy promotion and use** renewable, energy, res, electricity, support, sources, production, wind, grid, biofuels
18	**Member state climate policy assessment** energy, measures, government, efficiency, tax, policy, ministry, national, sector, electricity	18	**Mainstreaming climate policy into EU structural funds** climate, change, action, support, programme, programmes, development, actions, management, specific
19	**Improving the ETS** eu, emissions, cost, option, ets, cost, options, market, impact, carbon	19	**Revision of the ETS** ets, carbon, eu, allowances, allocation, emissions, free, costs, installations, revenues
20	**Future emissions reduction scenarios** scenario, costs, cost, baseline, analysis, scenarios, potential, impact, model, reduction	20	**Auctioning of emissions allowances** volume, allowances, price, average, total, market, period, trading, month, successful
21	**Electricity generation** electricity, power, sources, plants, production, renewable, plant, mw, energy, capacity	21	**Biomass and biofuel use and production** production, biomass, emissions, eu, land, ghg, carbon, biofuels, energy, supply
22	**Sustainable land-use** biomass, heat, production, land, energy, heating, wood, waste, potential, fuel	22	**Sustainable land-use** land, mitigation, agricultural, emissions, policy, management, potential, agriculture, action, cap
23	**External climate support** climate, change, development, policy, action, support, carbon, research, management, sustainable	23	**Financing Projects of Common Interest** energy, project, projects, european, investment, development, financial, investments, finance, eu
24	**Energy regulation and directives** article, directive, commission, european, ec, union, regulation, accordance, referred, official	24	**Implementing regulation on energy issues** values, test, regulation, power, fuel, speed, input, european, type, consumption
25	**Energy measurement vocabulary** kwh, water, gwh, kg, plan, mw, power, total, temperature, air	25	**Energy labelling and eco-design for products** energy, products, total, product, air, water, space, efficiency, heat, heating
26	**Monitoring and verification** data, monitoring, verification, reporting, section, operator, report, guidance, activities, eu	26	**Household energy prices and costs** energy, data, eu, consumption, countries, households, share, country, european, sector
27	**Modelling methodologies** data, directive, model, policy, level report, sector, approach, factors, results	27	**Economic analysis of energy policies and technologies** cost, costs, scenario, model, scenarios, analysis, energy, impact, technology, results
28	**Roadmap vocabulary** eu, transport, ghg, carbon, ets, scenario, reference, policies, growth, energy	28	**Scenarios for transport energy use and emissions** energy, transport, electricity, gas emissions, oil, generation, res, total, gross
29	**Common energy terminology** energy, transport, gas, oil, gross, electricity, generation, nuclear, indicators, consumption	29	**Competitiveness of transport sector** eu, costs, market, evaluation, cost, impacts, study, impact, legislation, competitiveness
30	**Project language** project, projects, development, financial, funding, capacity, support, eur, financing, implementation	30	**Emissions monitoring, verification and accreditation from maritime transport** information, relevant, system, data, requirements, article, monitoring, operator, compliance