Do bridges *dream* of water pollutants? Towards DreamsKG, a knowledge graph to make digital access for sustainable environmental assessment come true

Darío Garigliotti[1], Johannes Bjerva[1], Finn Årup Nielsen[2], Annika Butzbach[1], Ivar Lyhne[1], Lone Kørnøv[1], Katja Hose[1,3]

[1] {dariog@cs, jbjerva@cs, annikab@plan, lyhne@plan, lonek@plan, khose@cs}.aau.dk, Aalborg University, Denmark
[2] faan@dtu.dk, DTU, Denmark
[3] katja.hose@tuwien.ac.at, TU Wien, Austria

ABSTRACT

An environmental assessment (EA) report describes and assesses the environmental impact of a series of activities involved in the development of a project. As such, EA is a key tool for sustainability. Improving information access to EA reporting is a billion-euro untapped business opportunity to build an engaging, efficient digital experience for EA. We aim to become a landmark initiative in making this experience come true, by transforming the traditional manual assessment of numerous heterogeneous reports by experts into a computer-assisted approach. Specifically, a knowledge graph that represents and stores facts about EA practice allows for what it is so far only accessible manually to become machine-readable, and by this, to enable downstream information access services. This paper describes the ongoing process of building DreamsKG, a knowledge graph that stores relevant data- and expert-driven EA reporting and practicing in Denmark. Representation of cause-effect relations in EA and integration of Sustainable Developmental Goals (SDGs) are among its prominent features.

KEYWORDS

Knowledge graphs, Knowledge graph construction, Sustainability, Cause-effect relations

ACM Reference Format:
Darío Garigliotti[1], Johannes Bjerva[1], Finn Årup Nielsen[2], Annika Butzbach[1], Ivar Lyhne[1], Lone Kørnøv[1], Katja Hose[1,3]. 2023. Do bridges *dream* of water pollutants? Towards DreamsKG, a knowledge graph to make digital access for sustainable environmental assessment come true . In *Companion Proceedings of the ACM Web Conference 2023 (WWW '23 Companion), April 30-May 4, 2023, Austin, TX, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3543873.3587590

1 INTRODUCTION

The interest for making digital access a reality for large corpora of documents, abundant in textual content within many organizations and sectors, has extended the application cases of Information Extraction (IE) technologies to a variety of domains with heterogeneous data formats, including legal documents [22], gastronomy [23], financial reports [3, 20], and public crisis response [10]. More recently, IE applications have embraced modalities beyond text [8, 16] and within scenarios where scarcity meets data-hungry methods [20]. Environmental assessment (EA), although involving this kind of data phenomena, is still dominated by traditional practices, heavy on tedious, expert human labor. Large volumes of technical reports, highly heterogeneous and rich in EA content, make this area one of enormous potential for developing digital access to this vast amount of information. And being EA a legal requirement in most countries world-wide, developing such a sustainable digital transformation can have very large social and economical impact. Beyond a very few works in this area [6, 19], the lack of knowledge resources available is a key challenge to enable building digital access experiences for sustainable EA.

DREAMS[1] is an interdisciplinary project aiming to provide digital support for environmental assessment. In the context of this project, at the core of powering the future of digital access experience for EA, we place DreamsKG, a knowledge graph intended to represent and store facts about EA practice so far only accessible manually in the mostly textual content of numerous heterogeneous reports. We aim to tackle the challenge of building such a key knowledge resource in a domain where this kind of resources are very scarce. By building DreamsKG, we have at hand a resource that can power digital information access services that come to transform the traditionally manual practices of EA professionals. In particular, DreamsKG will contain causal relations between activities relevant to a variety of environment-sensitive projects, their effects and recipients, their significances, and the possible mitigations. Figure 1 shows an excerpt of the envisioned DreamsKG. These relations are often complex and the available data is very scarce to build automatic extraction approaches, hence the challenge that building our knowledge graph implies. DreamsKG also integrates Sustainable Developmental Goals (SDGs) in a localized manner within the Danish EA context. This paper describes the ongoing work on building DreamsKG, as well as discusses insights and challenges found throughout its progress. Specifically, our process starts with a conceptualization of the fundamental entity classes and relations to represent EA knowledge. An annotation experiment follows, in order to collect high-quality, human-labeled data that can be properly structured in a third phase, KG construction. We also present

WWW '23 Companion, April 30-May 4, 2023, Austin, TX, USA

ACM ISBN 978-1-4503-9419-2/23/04…$15.00
https://doi.org/10.1145/3543873.3587590

[1] https://dreamsproject.dk/

an ensemble of envisioned opportunities for research problems that we aim to tackle in order to enrich the capabilities of DreamsKG.

The paper is structured as follows. Section 2 introduces the problem setting around key initial resources. Section 3 describes our conceptualization of EA for the KG, and describes the annotation experiment to identify their instances in the reports. The ongoing DreamsKG construction is described in Section 4, whereas Section 5 complements it with our perspectives on future capabilities that DreamsKG could support. Section 6 concludes our work.

2 PROBLEM FORMULATION

The starting point in our process towards building DreamsKG is a scenario where few items are available: a corpus of environmental assessment (EA) reports and a compiled vocabulary of known relevant terms. This scenario is closely similar to the actual work setting that most of environmental assessors are presented with in their traditional practice. Our ambition is to transform the way that EA is carried out, enabling digital access to the relevant information. This goal requires transforming the corpus of reports into relevant machine-readable knowledge items that can be systematically retrieved and verified against information sources of the specific context that EA practitioners are working on.

2.1 Environmental assessment reports

An EA report is a document that mainly describes and assesses the environmental impact of a series of activities involved in the development of a project or plan in a particular site context. In it, each activity is considered throughout its entire life span, since planning, through its construction, during its operation (for example, the period by which a building or a power station is functional and active), all until its full decommission and removal. Some EA reports might contain further assessments regarding considered measures to mitigate the effects of an activity in a recipient within the environment. For example, for an activity of turbine wings rotating, with impact on bat populations, a mitigation measure could be regulating the operation hours of the turbine. The report is handled by the relevant authority of competence, as part of the legal procedure that regulates a project development proposal.

Even though making use of this kind of document is part of the standard practice in EA, the documents themselves are mostly not standard. Beyond some commonly present aspects described in these reports, they are instead usually heterogeneous in their content, extension, structure, and format. In the particular case of our corpus of EA reports, they are formatted in Portable Document Format (PDF). They have been created by exporting corresponding original documents created in commercial text processors, originals that are not available.

An EA report usually also describes key information about the kind of project or plan that it corresponds to, for example, a report about a project in the context of gas pipelines, or local electric grid, or solar panels, and so on.

The corpus we have access to in the DREAMS project collects around 2,100 EA reports developed in Denmark, written in Danish. This adds to the complexity of the envisioned scenario of extracting information from the reports, automatically, by relying on data-powered method, as this data is rather scarce for this language in contrast with more prominent languages.

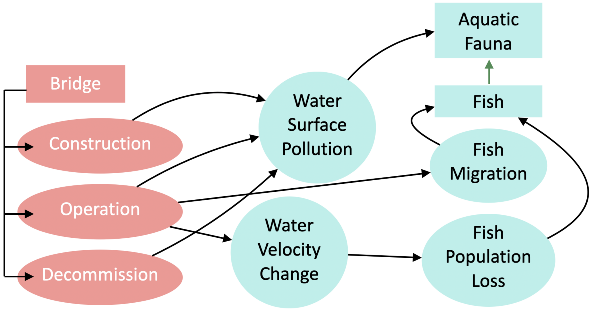

Figure 1: An excerpt of the envisioned DreamsKG. In red, at the left, an activity centered around a bridge and its several phases. In light blue, at the center and right, impacts (effects and recipients) caused by the activity in a given phase or by other impacts.

2.2 Vocabularies

Environmental assessors are equipped with a living vocabulary of terms relevant to their assessment practice. They search for a term of interest in an EA report to discover, via its occurrences, larger textual excerpts supporting and detailing the term, while also possibly containing other relevant information items.

In our work, we have at hand three vocabularies, each containing known impacts within one of three environmental parameters: biodiversity, climate, and health. These parameters have been selected due to being of strategic value: they are topical, diverse, and likely rich in content. The vocabularies are compiled by experts on each environmental parameter, and are structured in a shallow manner. The hierarchy comprises category, sub-category, and term, as in the examples from Table 1. The set of terms for a sub-category often presents a mixture of typical and topical information, the latter corresponding to a bag or cluster of topically related terms. The knowledge graph construction aims, among other goals, to accordingly break down these topicalities and retain only typical ontological subclass relations.

The vocabularies contain these hierarchy volumes:

- Climate: 9 sub-categories within 2 categories.
- Biodiversity: 15 sub-categories within 3 categories (Protected species, Habitats, and Ecosystems).
- Health: 19 sub-categories within 2 categories (Physical and environmental determinants, and Social determinants).

We employ the vocabularies as basic resource which to build the knowledge graph onto. As we discuss in the next section, the vocabularies are expanded via human annotation with new entries, and eventually serve as solid starting sets of entities for the knowledge graph.

3 CONCEPTUALIZATION AND ANNOTATION

Once we establish the basic elements in our problem scenario, as described in the previous section, we proceed to design and build DreamsKG. Firstly, we identify fundamental concept classes and

Table 1: Excerpt of the seed vocabulary for climate environmental parameter.

Category	Sub-category	Term
Climate impact	Direct greenhouse gas emissions	Greenhouse gas, emission, carbon dioxide, CO2e, CO2-eq, CO2eq, CO2, N2O
	Indirect greenhouse gas emissions	
Climate change	Heat waves	Temperature, heat wave, heat island, cooling
	Drying	Drying, evaporation, drying out
	Extreme precipitation and floods	Precipitation, weather events, showers, storm surges, cloudbursts, thaws

relations for representing EA knowledge. This conceptualization allows us to have a backbone structure with respect to which refine our considerations accordingly and populate the knowledge graph.

In a second stage, we design, set up and conduct an annotation experiment to collect high-quality, human-labeled data. The purpose of the annotation phase is manifold. Its objectives are not restricted to provide training data to power supervised machine learning of models for eventual relevant tasks at hand, such as recognizing and disambiguating entities. Rather, firstly, annotation proves to enhance the vocabularies, via the annotators inputting alternative aliases for the known entities as well as extending such a set with newly considered elements for the different classes. These expert annotations will become the fundamental resource to populate DreamsKG. Additionally, by capturing the correspondence of a knowledge item with the textual passage where it is mentioned, DreamsKG can also represent evidence from EA reports for its facts. And as we observe throughout this annotation stage, it overall serves for the refining updates on the conceptualization of the atomic components of DreamsKG, namely, the actual set of entity classes and relationships to be represented.

3.1 Conceptualization

DreamsKG aims to address a central challenge in our problem, this is, to provide information access to valuable cause-effect relations among entities in environmental assessment. A paradigmatic triple in our KG, hence, should capture the relation from an environmental phenomenon leading to a consequence. In light of these consideration, the fundamental entity classes to represent EA knowledge are:

- *Activity*: a human work, of sufficient relevance regarding its environmental impact, involved in the development of a project or plan.
- *Impact*: a consequence, of sufficient environmental relevance, that follows a phenomenon (e.g. an activity, or another impact).
- *Mitigation measure*: a process that is intended to mitigate the consequences of an environmental phenomenon.
- *Sustainable Developmental Goal (SDG)*: an overarching goal for a desired outcome within a specific area of human development.

Ideally, these fundamental cause-effect relations would be captured between those classes:

- An activity causes an impact.
- A mitigation measure minimises an impact.
- An impact affects an SDG.
- A mitigation measure may reduce a negative impact on an SDG.

Additionally, these attributes are fundamental in our conceptualization of EA:

- The *phase* in which an activity is: Planning, Construction, Operation, or Decommission.
- The *significance* of an impact, this is, the polarity (positive or negative) and degree of magnitude by which a relevant phenomenon affects the environment. Impact significance is a central concept in EA that is defined in various ways [14, 15, 17].

As an example relating these concepts, consider an EA report describing a project centered on wind turbines. A phrase "Rotating wings pose a serious risk to bats" in the report informs about the activity of turbine wings rotating in the operation of the turbine (or some aspect about this phenomenon, like the sound from this rotation, that should be clearer from the textual context in the report) leads to risk of collision, an impact that affects bats. A complementary phrase "regulation of operation hours of the turbine" could be identified as a mitigation measure for the impact of this activity.

As we describe later in this article, although rather intuitive, our conceptualization may be involving assumptions that are idealized, or with a suboptimal granularity (i.e. too coarse, or too fine). Hence operationalizing these fundamental entity classes and relations requires design decisions for adjusting it to the actual occurrences of these items in the EA reports. Some of those designs were adjusted during the setting up of the annotation, while some others emerge through the iterations of the running annotation experiment.

3.2 Annotation platform

The corpus of EA reports presents a series of challenges when we need to carry out this annotation experiment. As we mentioned, the reports are very heterogeneous in the format and the structure of their content. Moreover, they are available in Portable Document Format (PDF) file format, and we do not have access to the respective original documents editable in commercial word processor software. The nature of this EA corpus makes it difficult to establish heuristics for processing them optimally so that their content can be annotated. After experimenting with several software tools for PDF-to-text content extraction, we still found minor mistakes across several reports, as well as recurrent issues with properly processing page numbering, headers, footers, and others. Even with these extraction issues ideally solved, we would still be in presence of report content in plain text, not necessarily comfortable to be annotated by humans.

We finally conducted the annotations on tagtog [2] [2]. Tagtog is a tool where invited annotators can work on a project accessible online via web browser. On Tagtog, multiple workers can annotate on a PDF-like representation of the EA report, while an underlying correspondence with its plain text content in automatically determined sections of the PDF is also accessible. Features like merging annotations, automatic pre-annotation from importable

[2] https://www.tagtog.com/

Table 2: Details and statistics of the annotation experiment outcome.

Plan or project type	Details	Number of reports	Number of annotations
Rail infrastructure projects	Include new rail routes as well as electrification of existing rail routes	9	644
Energy infrastructure projects	Include energy and gas infrastructure	6	796
Road infrastructure projects	Include highways as well as high classified roads	16	1,795
Municipal plans	Within the last five years as a key planning type in Denmark	64	1,161
Photovoltaic projects		17	402
Total	-	112	4,798

dictionary of surface forms, and in-house training of supervised machine learning complement the neat visualization of an ongoing annotation. Relevant to our labeling experiment is the possibility to define several entity classes and relations to be annotated, as we indeed have in our conceptualization, and the ability to extend entries of imported dictionaries with newly discovered entity aliases.

3.3 Annotation task

Our annotation experiment was expected to encompass all the entity classes and relations that we determined in our conceptualization of EA. During the setting up of the annotation experiment, we revisit the concepts previously included within the annotation scope. In order to actually make them operational towards the construction of DreamsKG, we make these considerations in the design of our annotation task:

- The annotation of impacts would often involve making the correspondence to impacts known from the initial vocabulary. The activities, instead, are to be found anew.
- In the case of annotating a relation between an activity and an impact, we also annotate the verbal phrase that links them in natural language as the *predicate* of the relation. This textual expression for the predicate is deemed as useful when eventually performing relation extraction.
- The annotation of the significance of an impact is unfolded into annotating both (i) a categorical label for its polarity (positive or negative) and degree (significant, highly significant), and (ii) the textual expression, if any, that corresponds to the significance.
- Annotating a mitigation measure reduces to annotate a possibly long proposition within a textual sentence, or one or more sentences, that describe such a mitigation measure. This very likely leads to a sparse set of very unique phrases, for describing mitigation processes that are rather already scarce through the reports. A shorter textual unit, as done for activity or impact, would in most cases fail to capture such a mitigation process completely.
- SDGs, as well as their relations with other classes in our conceptualization, are not to be annotated, but rather dealt with separately by their expert-driven representation and integration in DreamsKG.

Our pool of annotators consists of experts in EA as well as M.Sc. students in the field, who interact with the tagtog platform by reading through an EA report and following our annotation guidelines. We understand that systematically considering all the interplaying classes and relations would lead to a better annotation outcome. Given also the limited available human resources for the annotation, we avoid separated annotation experiments, one per concept, or separated by classes versus relations, etc. Instead, we design the annotation experiment in a way so that each expert annotates everything that is relevant as she reads through the report.

In the Tagtog annotation platform, we model these designs and deploy guidelines for the annotators. We also import the vocabularies as dictionaries, one per environmental parameter, mapping a designated unique identifier of an impact to its known surface forms.

Once the set up of the annotation experiment is completed, we proceed to conduct the annotation work. This work is actually carried out in several iterations. In early iterations, we experiment with outcome of small annotation instances, as well as by considering the feedback from the annotators regarding the difficulty of their work, and the suggestions they provide about the phenomena they observe. Upon these resulting status, we refine the designs accordingly in our conceptualization and in the actual annotation platform. Throughout these annotation iterations, we come to represent a number of patterns present as phenomena between entity classes and relations. The prominent patterns are depicted in Fig. 2, including this kind of phenomena:

- Two or more activities (related within the context described in the EA report) lead to a common impact. For some of them, it is also possible to determine an explicit textual phrase for its predicate.
- Symmetrically, an activity leads to two or more impacts (with similar possibilities regarding presence of textual expression for the predicate).
- Two closely related activities preliminarily identified as such in separation, may rather be assumed to encompass a single activity, where each of the two parts originally identified does not provide sufficient detail only on itself. For example, parts about "gravel transport" and "construction of a bridge" may better refer together to the activity of gravel transport related to construction of a bridge (and not related to others in the project).
- Two impacts can be represented as causally related, as one is consequence of the other. For example, the pollution of a body of water leading to the migration of an animal population living in or near by such a body of water.

Table 2 presents the outcome of our annotation experiment, over a selected set of types of projects or plans deemed as high-priority for DREAMS. The number of reports refers to the total of EA reports thoroughly annotated per project or plan type. The last column counts total of complete annotations in all those reports across the two annotators per report, where an annotation is considered complete if (i) it identifies an instance of the fundamental causal

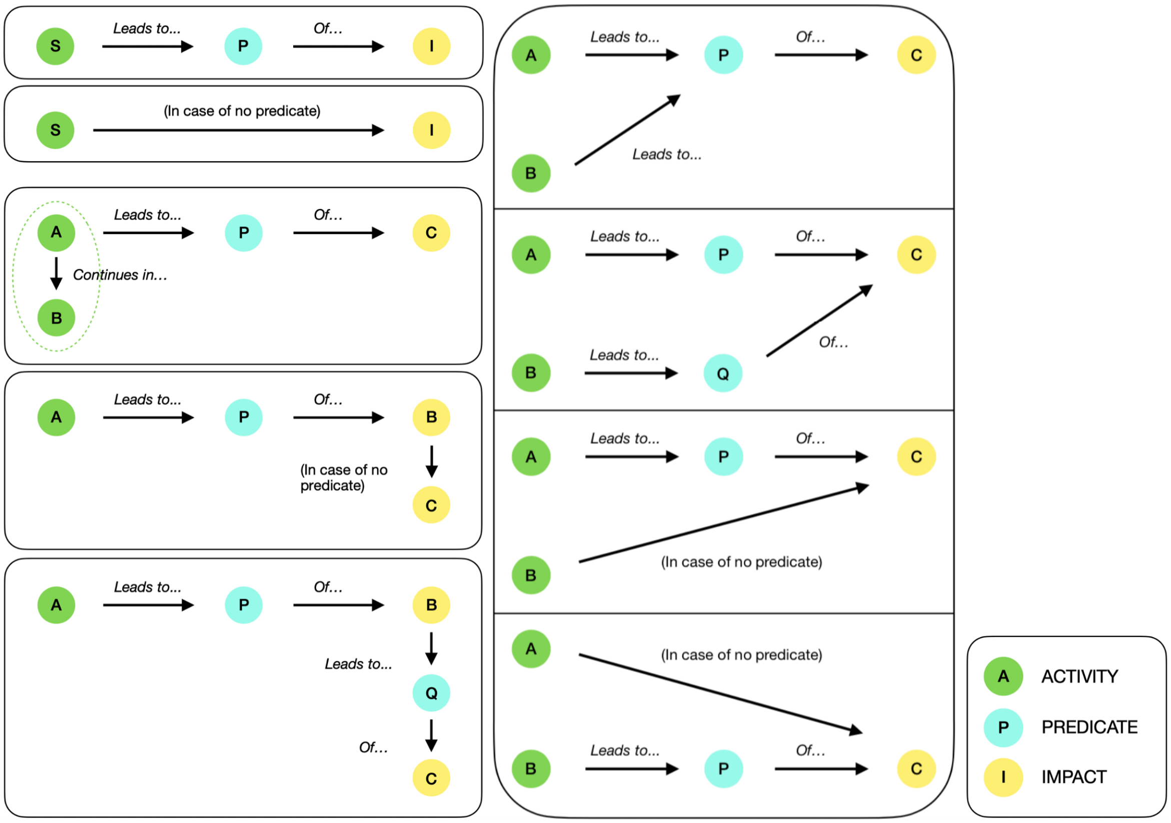

Figure 2: Causal relation patterns to be captured in the annotation task, involving activities (in green), predicates (in light blue) and impacts (in yellow). In each box, we depict causal order in the paradigmatic relation activity - (predicate) - impact.

relation of interest, activity leading to impact, and also includes (ii) annotated impact significance and (iii) annotated activity phase.

As a measure of quality of the annotation experiment, we report on the inter-annotator agreement (IAA) that is available on tagtog separately for each annotated class. The average pairwise IAA is rather low: 22.88% for annotating activities and 27.38% for impacts. A main consequence of this finding is that further processing towards DreamsKG via automatic extraction methods is largely hindered, and a strong element of manual quality control of the facts to be part of DreamsKG seems pertinent.

4 KG CONSTRUCTION

The ongoing construction of our knowledge graph is undertaken as an iterative process of refinements, where researchers in EA and computer science team up to provide structure to the annotated data in line with our conceptualization. A thorough manual verification of the triples is performed, in particular, via disambiguation and unification across entity surface forms. Through reification, a basic triple for the paradigmatic relation of an activity leading to an impact is further coupled with other entities like second-order impacts or mitigation measures. A distinguished entity class of importance is the paragraph (uniquely identified automatically by the annotation platform) of the report from which a phenomenon is extracted as a triple. This entity class is added to our conceptualization, as it represents the source of evidence supporting a statement claimed by DreamsKG.

This iterative KG building often involves stakeholders participating in the DREAMS project, from whom the researchers elicit valuable insights and qualifications. These exchanges with the professionals in the EA practice has already led to relevant refinements:

- The simplifying concept of impact gets unfolded into an effect and its recipient. This follows a more natural design, and solves some heterogeneities present through the annotated data.
- Such an unfolding as described in the previous item uncovers implicit effects missing in the annotation, possibly very difficult to determine even for experts. As an example, consider the activity of building a bridge, and its identified impact in fish populations nearby. after the unfolding of impact concept, the fish becomes the recipient of a to-be-determined effect, that could be migration, death, change in food availability, or others.
- The project or plan type directing the EA report is learnt to be highly relevant for the practitioners. Indeed, it is very early in the EA workflow that project type is considered as first decision node from which to continue in seeking to identify the main classes and relations. This priority in the workflow could be reflected in the ontological structure underlying DreamsKG.

Domain experts participating in the DREAMS project have separately developed relevant knowledge graphs for the rest of the environmental parameters, this is, climate and health, due to lack of data in the report corpus found upon inspection of a large sample of reports. The ongoing DreamsKG construction process also involves performing an alignment between the graphs obtained for the three parameters: biodiversity, climate, and health.

Furthermore, coupled activity and impact information is being manually mapped to SDG targets. This SDG integration builds on the work of other researchers participating in DREAMS, who have developed a framework for sustainable developmental targets (subcategories of goals, more operational than its parent categories) [1]. The integration, appropriately, performs localization of SDGs, i.e., takes into account the particular Danish context and the EA context. In this way, we incorporate neatly the class from our EA conceptualization that was left out during the annotation experiment.

A number of additional aspects that emerge through these data structuring processes are also in consideration:

- Certain heterogeneity from the more authoritative process by which the KGs for other environmental parameters are built.
- The heterogeneity between the passage-related entity class to represent textual evidence, and the EA-related concepts.
- Achieving balance across expectations and interests of different stakeholders that could affect the proper development of the DreamsKG construction.

5 PERSPECTIVES ON OPPORTUNITIES

We recognize that the information space around DreamsKG lends itself to a variety of further capabilities to be explored. In this section, we give a thought to a handful of features that would enhance the digital experience in EA and the business opportunities that DreamsKG aims to power.

5.1 Overcoming limited data

Within the family of methods typically employed for Information Extraction (IE) tasks, some data-intensive approaches require a corresponding abundance of data that matches the kind of the new data where to extract from. The space of resources associated to the EA data behind DreamsKG is limited due to at least a couple of factors. On the one hand, most of the data used by such methods is in English, while reports in the corpus available in our DREAMS project are in Danish language. On the other hand, the EA domain is essentially more restricted than the general-domain data collections on which many methods rely.

In particular, given the international extent where DREAMS is framed in, the limitations regarding resources available in a language would become even more pronounced once DreamsKG is profiled for dealing with knowledge in other languages. Nevertheless, overcoming these limitations would strongly contribute to DreamsKG consolidating as a landmark in further stages of EA digitalization. Hence the importance of strategies for building robustly when structuring across languages.

5.2 Predicting impact significance

Impact significance is a central concept in EA, and has long been a topic of debate in the field, in part due to the difficulty in defining it [14, 15, 17, 21, 24]. Among the different definitions of significance that lead to such a difficulty, some stem from the process in which significance is determined [25], while some from the individuals determining the significance itself [21]. In recent years, there has been some push towards concretising and simplifying this process [4, 5, 7].

To corroborate this kind of phenomena, we conduct an experiment where an EA researcher compiles 100 instances of annotated impacts. The selection is not strictly required to be at random, only encouraged, since more than verifying that identifying significance is hard (which we do), we are also interested in compiling a variety of main group of cases showing why it is hard. We find indeed diverse situations, such as the frequent absence of explicit terms about significance, or the need to decide on significance within a past context clearly beyond the instance phrase of interest. We observe that when significance can be determined, is mostly negative.

Although approaches from recent advances in representation learning are, a priori, promising, they are also affected by the domain and language limitations previously described in Section 5.1. The problem of sentiment analysis [18, 26, 27], where some textual unit is determined as positive or negative in sentiment, appears to be the closest area of research in language technology, and it is from it that we would start experimenting on significance prediction.

5.3 Advancing in SDG awareness

With increasing amount of recent research on linking SDGs to EA [1, 9, 11, 12], a starting opportunity around SDG is the possible development of a computational approach to automatically align SDG targets with fundamental concepts within DreamsKG.

Beyond this initial purpose, we identify a more ambitious problem space in the SDG awareness: not only determining which SDG (targets) an impact is related to, but also to what degree it is integrated. Kørnøv et al. [13] implement a spectrum containing six levels of SDG and EA integration, which scale ranges from SDG-washing in EA to SDG-led EA. In this setting, it would be relevant to investigate the degree to which EA information instances are using SDGs in a meaningful way, or are at risk of SDG-washing.

6 CONCLUSION

In this paper, we present our ongoing efforts in constructing DreamsKG, a knowledge graph that aims to represent cause-effect relations between relevant concepts of environmental assessment, and that we deem key to enable the digital transformation into a sustainable EA experience.

We consider that the discussed insights and challenges found throughout the KG construction, and the capabilities further envisioned, would contribute to enhance our whole DREAMS project to be centered around DreamsKG, to project itself beyond the Danish EA context into a landmark that truly came true.

Funding: This research was partially funded by the Innovation Fund Denmark (Innovationsfonden) (grant agreement number, 0177-00021B DREAMS), the Poul Due Jensen Foundation, and the Danish Council for Independent Research (DFF) under grant agreement no. DFF-8048-00051B.

Acknowledgments: This research is part of the DREAMS project, Digitally Supporting Environmental Assessment for the Sustainable Development Goals. The authors gratefully acknowledge the contributions of Kasper David Pedersen Storrs throughout the annotation experiment, as well as the ongoing contributions of Emilia Ravn Bøss on the SDG framing and integration, and Gabriel Gulis on the development of health-related models.

REFERENCES

[1] Emilia Ravn Bøss, Lone Kørnøv, Ivar Lyhne, and Maria Rosario Partidario. 2021. Integrating SDGs in environmental assessment: Unfolding SDG functions in emerging practices. *Environmental Impact Assessment Review* 90 (2021), 106632.

[2] Juan Miguel Cejuela, Peter McQuilton, Laura Ponting, Steven Marygold, Raymund Stefancsik, Gillian Millburn, and Burkhard Rost. 2014. tagtog: interactive and text-mining-assisted annotation of gene mentions in PLOS full-text articles. *Database: the journal of biological databases and curation* 2014 (01 2014), bau033. https://doi.org/10.1093/database/bau033

[3] Pan Ding, Liang Zhuoqian, and Deng Yuan. 2020. Textual Information Extraction Model of Financial Reports. In *Proceedings of the 2019 7th International Conference on Information Technology: IoT and Smart City* (Shanghai, China) *(ICIT 2019)*. Association for Computing Machinery, New York, NY, USA, 404–408. https://doi.org/10.1145/3377170.3377231

[4] Carla Grigoletto Duarte and Luis Enrique Sanchez. 2020. Addressing significant impacts coherently in environmental impact statements. *Environmental Impact Assessment Review* 82 (2020), 106373.

[5] Álvaro Enríquez-de Salamanca. 2021. Simplified environmental impact assessment processes: Review and implementation proposals. *Environmental Impact Assessment Review* 90 (2021), 106640.

[6] Julián Garrido and Ignacio Requena. 2011. Proposal of ontology for environmental impact assessment: An application with knowledge mobilization. *Expert Systems with Applications* 38, 3 (2011), 2462–2472. https://doi.org/10.1016/j.eswa.2010.08.035

[7] Gesa Geißler, Johann Köppel, and Marie Grimm. 2022. The European Union Environmental Impact Assessment Directive: Strengths and weaknesses of current practice. In *Routledge Handbook of Environmental Impact Assessment*. Routledge, 282–301.

[8] Dihong Gong, Daisy Zhe Wang, and Yang Peng. 2017. Multimodal Learning for Web Information Extraction. In *Proceedings of the 25th ACM International Conference on Multimedia* (Mountain View, California, USA) *(MM '17)*. Association for Computing Machinery, New York, NY, USA, 288–296. https://doi.org/10.1145/3123266.3123296

[9] Theo Hacking. 2019. The SDGs and the sustainability assessment of private-sector projects: Theoretical conceptualisation and comparison with current practice using the case study of the Asian Development Bank. *Impact Assessment and Project Appraisal* 37, 1 (2019), 2–16.

[10] Roberto Interdonato, Antoine Doucet, and Jean-Loup Guillaume. 2020. Unsupervised Crisis Information Extraction from Twitter Data. In *Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (Barcelona, Spain) *(ASONAM '18)*. IEEE Press, 579–580.

[11] Olga Kolesnichenko, Lev Mazelis, Alexander Sotnik, Dariya Yakovleva, Sergey Amelkin, Ivan Grigorevsky, and Yuriy Kolesnichenko. 2021. Sociological modeling of smart city with the implementation of UN sustainable development goals. *Sustainability Science* 16, 2 (2021), 581–599.

[12] Lone Kørnøv. 2021. SEA as a change agent: Still relevant and how to stay relevant? *Impact Assessment and Project Appraisal* 39, 1 (2021), 63–66.

[13] Lone Kørnøv, Ivar Lyhne, and Juanita Gallego Davila. 2020. Linking the UN SDGs and environmental assessment: Towards a conceptual framework. *Environmental Impact Assessment Review* 85 (2020), 106463.

[14] David P Lawrence. 2007. Impact significance determination—back to basics. *Environmental Impact Assessment Review* 27, 8 (2007), 755–769.

[15] David P Lawrence. 2007. Impact significance determination—pushing the boundaries. *Environmental Impact Assessment Review* 27, 8 (2007), 770–788.

[16] Yunfeng Li, Peijun Du, Zhaohui Xue, Le Gan, Xin Wang, and Hao Liang. 2018. A Method of Rice Information Extraction Based on Particle Swarm Optimization SVM Algorithm. In *Proceedings of the 2018 10th International Conference on Machine Learning and Computing* (Macau, China) *(ICMLC 2018)*. Association for Computing Machinery, New York, NY, USA, 396–400. https://doi.org/10.1145/3195106.3195155

[17] Ivar Lyhne and Lone Kørnøv. 2013. How do we make sense of significance? Indications and reflections on an experiment. *Impact Assessment and Project Appraisal* 31, 3 (2013), 180–189.

[18] Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. 2011. Learning Word Vectors for Sentiment Analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, Portland, Oregon, USA, 142–150. https://aclanthology.org/P11-1015

[19] Edrisi Muñoz, Elisabet Capón-García, José M. Laínez, Antonio Espuña, and Luis Puigjaner. 2013. Considering environmental assessment in an ontological framework for enterprise sustainability. *Journal of Cleaner Production* 47 (2013), 149–164. https://doi.org/10.1016/j.jclepro.2012.11.032 Cleaner Production: initiatives and challenges for a sustainable world.

[20] Minh-Tien Nguyen, Dung Tien Le, Le Thai Linh, Nguyen Hong Son, Do Hoang Thai Duong, Bui Cong Minh, Nguyen Hai Phong, and Nguyen Huu Hiep. 2020. AURORA: An Information Extraction System of Domain-Specific Business Documents with Limited Data. In *Proceedings of the 29th ACM International Conference on Information &amp; Knowledge Management* (Virtual Event, Ireland) *(CIKM '20)*. Association for Computing Machinery, New York, NY, USA, 3437–3440. https://doi.org/10.1145/3340531.3417434

[21] Kaja Peterson. 2010. Quality of environmental impact statements and variability of scrutiny by reviewers. *Environmental Impact Assessment Review* 30, 3 (2010), 169–176.

[22] Verónica Romero, Alicia Fornés, Emilio Granell, Enrique Vidal, and Joan Andreu Sánchez. 2019. Information Extraction in Handwritten Marriage Licenses Books. In *Proceedings of the 5th International Workshop on Historical Document Imaging and Processing* (Sydney, NSW, Australia) *(HIP '19)*. Association for Computing Machinery, New York, NY, USA, 66–71. https://doi.org/10.1145/3352631.3352637

[23] Nuno Silva, David Ribeiro, and Liliana Ferreira. 2019. Information Extraction from Unstructured Recipe Data. In *Proceedings of the 2019 5th International Conference on Computer and Technology Applications* (Istanbul, Turkey) *(ICCTA '19)*. Association for Computing Machinery, New York, NY, USA, 165–168. https://doi.org/10.1145/3323933.3324084

[24] Riki Therivel. 2004. *Strategic environmental assessment in action*. London: Earthscan.

[25] Graham Wood and Julia Becker. 2004. Evaluating and communicating impact significance in EIA: A fuzzy set approach to articulating stakeholder perspectives. In *Presentation to the international association of impact assessment conference, Vancouver, Canada*. 26–29.

[26] Hu Xu, Bing Liu, Lei Shu, and Philip Yu. 2020. DomBERT: Domain-oriented Language Model for Aspect-based Sentiment Analysis. In *Findings of the Association for Computational Linguistics: EMNLP 2020*. Association for Computational Linguistics, Online, 1725–1731. https://doi.org/10.18653/v1/2020.findings-emnlp.156

[27] Wei Xue and Tao Li. 2018. Aspect Based Sentiment Analysis with Gated Convolutional Networks. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Melbourne, Australia, 2514–2523. https://doi.org/10.18653/v1/P18-1234