

A Government Support System Based on Citizens' Interactions:

A Comparative Study of Different Arabic Language Analysis Methodologies

Submitted By: Mohamed Adel Rezk

A dissertation submitted in partial fulfilment of the requirements of
Alexandria University for the degree of
M.Sc. in Information Systems and Computers (Management Information
Systems)

June 2017

I certify that this dissertation which I now submit for examination for the award of MSc in Information Systems and Computers (Management Information Systems), is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

This dissertation was prepared according to the regulations for postgraduate study of the Alexandria University and has not been submitted in whole or part for an award in any other Institute or University.

The work reported on in this dissertation conforms to the principles and requirements of the Institute's guidelines for ethics in research.

Signed: _____

Date: **01 06 2017**

TABLE OF CONTENTS

TABLE OF CONTENTS.....	0
ABSTRACT.....	0
ACKNOWLEDGMENT.....	3
TABLE OF FIGURES	4
TABLE OF TABLES	6
TABLE OF EQUATIONS.....	6
1 Introduction.....	0
1.1 Research objectives.....	4
1.1 Research methodology	4
1.2 Scope and limitations	5
1.3 Organisation of the dissertation	5
2 Literature review	6
2.1 Background.....	6
2.2 Knowledge Management Process	7
2.3 Artificial Intelligence Tools.....	7
2.4 A Comparative Study of Different Arabic Language Analysis Methodologies	8
2.4.1 <i>Challenges of Arabic NLP</i>	8
2.4.2 <i>Current state-of-the-art in Arabic NLP Resources and Tools</i>	9
2.4.3 <i>Fitting Arabic Natural Language Processing into our Public Policy Management Framework</i>	9
2.5 Citizen to Government Relation Analysis	10
2.6 Text analysis accuracy measuring methodology.....	11
2.7 Conclusion.....	11
3 Solution Components.....	12
3.1 Core Public Policy Vocabulary.....	12
3.1.1 <i>Component Overview</i>	12
3.1.2 <i>Component Background and Related Work</i>	14
3.1.3 <i>Public Policy</i>	14
3.1.4 <i>Public Policies Types</i>	14
	0

3.1.5	<i>Public Policies Job</i>	15
3.1.6	<i>Vocabularies</i>	15
3.1.7	<i>Ontology</i>	16
3.1.8	<i>Related Work</i>	16
3.1.9	<i>Methodology</i>	22
3.1.10	<i>Vocabulary Design</i>	25
3.1.11	<i>Conceptual Model</i>	25
3.1.12	<i>URI Naming Patterns</i>	27
3.2	Public Policy Collaboration Mining for Citizen Satisfaction	28
3.2.1	<i>Component Overview</i>	28
3.2.2	<i>Component Background and Related Work</i>	29
3.2.3	<i>Open Government Data</i>	29
3.2.4	<i>Natural Language Processing</i>	30
3.2.5	<i>Named Entity Recognition</i>	30
3.2.6	<i>Natural Language Processing in E-Government</i>	31
3.2.7	<i>Mining for Collaboration</i>	31
3.3	GOV DAF.....	32
3.3.1	<i>Component Overview</i>	32
3.3.2	<i>Gov-DAF knowledge base building pipeline</i>	34
3.3.3	<i>Approach independent Gov-DAF knowledge base building pipeline stages</i>	35
3.4	Conclusion.....	38
4	Experimentation and Evaluation	39
4.1	Core Public Policy Vocabulary.....	39
4.1.1	<i>Validation</i>	39
4.1.2	<i>Competency Questions</i>	39
4.1.3	<i>Applications</i>	40
4.1.4	<i>Abstract Use Cases</i>	40
4.1.5	<i>Scenarios</i>	42
4.1.6	<i>Experiments</i>	43
4.1.7	<i>Protégé Implementation</i>	43
4.1.8	<i>Protégé Screen Samples</i>	43
4.1.9	<i>CKAN Implementation</i>	49
4.1.10	<i>CKAN Screen Samples</i>	50
4.2	Public Policy Collaboration Mining for Citizen Satisfaction	53
4.2.1	<i>Semantic Profiling for Collaboration Mining</i>	53

4.2.2	<i>Profiling the Catalogues</i>	54
4.2.3	<i>Interlinking Publishers</i>	56
4.3	GOV DAF.....	59
4.3.1	<i>Topic Modelling for Public Policy Analysis Module</i>	59
4.3.2	<i>Citizen Satisfaction Index Estimation Pipeline</i>	61
4.3.3	<i>Public Policy Data Extraction and Cleansing Module</i>	61
4.3.4	<i>Public Policy Topic Modelling Module</i>	63
4.3.5	<i>Public Policy Co-occurrence Matrix Building Module</i>	64
4.3.6	<i>Public Policy Related Tweets Collection Module</i>	66
4.3.7	<i>Public Policy Matrix Based Related Tweets Querying Module</i>	68
4.3.8	<i>Public Policy Sentiment Analysis Module</i>	69
4.3.9	<i>Building Prediction Model and Citizen Satisfaction Estimation Module</i>	71
4.4	Conclusion.....	71
5	Conclusion	72
5.1	Experimentation, Evaluation and Limitation.....	72
5.1.1	<i>Core Public Policy Vocabulary</i>	72
5.1.2	<i>Formal Adoption of the Core Public Policy Vocabulary</i>	72
5.1.3	<i>Public Policy Collaboration Mining for Citizen Satisfaction</i>	73
5.1.4	<i>Gov – DAF</i>	76
5.2	Future Work and Open Research Issues	77
5.2.1	<i>Core Public Policy Vocabulary</i>	77
5.2.2	<i>Public Policy Collaboration Mining for Citizen Satisfaction</i>	77
5.2.3	<i>Gov-DAF</i>	78
Bibliography	79
APPENDIX A	92
APPENDIX B	93
APPENDIX C	94

ABSTRACT

This research is to produce A Public Policy Management Framework that includes the following components: (1) A ready to use Decision Support System for Governmental Organizations by Using Natural Language Processing (NLP) Techniques, Machine Learning Algorithms (ML) and Web 0.2 Concepts and Technologies for Citizen to Government (C2G) interactions Insights and Analysis, i.e. by applying different NLP and Data Mining algorithms to extract knowledge and predict actions. This will lead to adding citizen satisfaction variable to the public policy decision-making equation [Gov-DAF]. (2) Extendable Public Policy Knowledgebase [CPPV] which allow fundamental Public Policy Storage and linking for a wide range of use cases including Public Policy learning, with the ability of extending this Knowledgebase for other use cases e.g. Citizen to Government Interaction Analysis [Gov-DAF]. (3) Collaboration Mining tool which one of its use cases is to enable Public Policy Collaboration suggestions based on organization semantic profiles retrieved, built, and compared within the tool.

To achieve the aforementioned goals we are proposing a Government Support Framework for Public Policy Management That analyse governmental public policies (Introduced or Under discussions) and produce citizens' satisfaction rates (Real-time and Predicted) towards the policy and its Aspects by using Natural Language Processing (NLP) Techniques and Knowledge Management Processes for Citizens' satisfaction..

The outcome citizens' satisfaction index anent public policies is a core political research question. Seeking agile and efficient public policies, canny public policy makers are perpetually investigating how to weigh citizens' satisfaction index towards their policies, in order to overhaul faulty public policy aspects or topics that outcome a negative citizens' satisfaction index. Multiple automated and manual citizen's satisfaction index calculation methods are applied [1]–[10], bearing in mind that a correct calculation will grant a significant success for the public policy and the public policymakers. Our perspective is that nevertheless the index was well calculated or not it is always too late, public policy is already issued and citizen reactions are ensue. Hence, we proposed a public policy satisfaction prediction framework [11], [12], this

framework reckon on a knowledge base that allow formulating the prediction formulas.

To acquire this knowledgebase we designed, and extended the Core Public Policy Vocabulary [12]–[16], apply Named Entity Recognition and Topic Modelling in parallel for keywords extraction and semantic similarity measuring for relating the detected keywords with the pre-defined public policy aspects. The aforementioned methods will allow an automated population of the prediction knowledge base.

To increase policy acceptance we developed a new policy collaboration tool that will enable a well analysed and developed public policies according to the increased governmental collaboration. Due to the increasing adoption of open data among governments worldwide especially in the European Union area, a deeper analysis of the newly published data is becoming a mandate. Apart from analyzing the published dataset itself we aimed on analyzing published dataset catalogues. A dataset catalogue or a dataset metadata contains features that describe what the data is about in a textual representation. So, we first acquire data from open data portals, choose descriptive dataset catalogue features, and then construct an aggregated textual representation of the datasets. Afterwards we enrich those textual representations using Natural Language Processing (NLP) methods to create a new comparable data feature “Named Entities”. By mining the new data feature we are able to produce datasets and publishers relatedness network. Those networks are used to point similarities between the published data across multiple open data portals. Pointing all possible collaborations for integrating and standardizing data features and types would increase the value of data and ease its analysis process.

Key words:

Government Decision Support; Decision Analytics; Bayesian Policy Acceptance Prediction; Citizen Satisfaction; Policy Aspects; Opinion Mining; Sentiment Analysis; Semantic Relatedness; Citizens’ Satisfaction Index; Core Public Policy Vocabulary; Public Policies; Policy Satisfaction Prediction; Citizen Reactions; Topic Detection; Semantic Similarity; Automated Ontology Population; Policy Aspects; Government Decision Support; Decision Analytics; Bayesian Policy Acceptance Prediction; Citizen Satisfaction; Policy Aspects; Opinion Mining; Sentiment Analysis ;Semantic Relatedness; Public policies; public satisfaction; Government to Government

Learning; Federated Policy Repository; Public Policy Analysis; Core Public Policy Vocabulary; Semantic Policy Search Engine; Government Learning; Predictive Policy Analytics; Arabic Natural Language Processing.

ACKNOWLEDGMENT

TABLE OF FIGURES

FIGURE 1 CITIZEN SATISFACTION ANALYSIS PIPELINE Gov-DAF.....	1
FIGURE 2 CORE PUBLIC POLICY VOCABULARY (CPPV) UML [13].....	2
EQUATION 1 Gov-DAF CITIZEN SATISFACTION FORMULAS ‘SIMPLIFIED’	3
FIGURE 3 Gov-DAF ABSTRACTED [11], [12].....	3
FIGURE 4 OPEN DATA CATALOGUES AND PUBLISHERS’ SEMANTIC PROFILING CONCEPTUAL FRAMEWORK.....	4
FIGURE 5 DESIGN SCIENCE RESEARCH METHODOLOGY.....	5
TABLE 1 ARABIC NATURAL LANGUAGE PROCESSING CHALLENGES	8
TABLE 2 CURRENT STATE-OF-THE-ART IN ARABIC NLP RESOURCES AND TOOLS.....	9
FIGURE 6 THE ROLE OF CONCEPTUAL MODEL IN SYSTEMS DEVELOPMENT [122].....	22
FIGURE 7 SYSTEMATIC LITERATURE REVIEW [130]	23
FIGURE 8 THE CORE PUBLIC POLICY VOCABULARY MODELLING METHODOLOGY.....	24
TABLE 3 CPPV CLASSES I.E. CONCEPTS	25
TABLE 4 CPPV OBJECT PROPERTIES I.E. RELATIONS	26
TABLE 5 URI NAMING PATTERNS.....	27
FIGURE 9 IRISH GOVERNMENT’S OPEN DATA PORTAL.....	30
FIGURE 10 CPPV POLICY ANALYTICS EXTENSION.....	34
FIGURE 11 Gov-DAF KNOWLEDGE BASE BUILDING PIPELINE USING NAMED ENTITY RECOGNITION BASED METHODOLOGY	36
TABLE 6 “SIMPLIFIED” REPRESENTATION OF GOV-DAF KNOWLEDGE BASE IN TURTLE	36
FIGURE 12 Gov-DAF KNOWLEDGE BASE BUILDING PIPELINE USING TOPIC MODELLING BASED METHODOLOGY	38
TABLE 7 COMPETENCY QUESTIONS.....	40
FIGURE 13 CORE PUBLIC POLICY VOCABULARY (CPPV) USE CASES	42
FIGURE 14 PROTÉGÉ IMPLEMENTATION (CPPV METADATA)	44
FIGURE 15 PROTÉGÉ IMPLEMENTATION (CPPV CLASSES, PROPERTIES AND DATA TYPES)	44
FIGURE 16 PROTÉGÉ IMPLEMENTATION (CPPV INDIVIDUALS)	45
FIGURE 17 PROTÉGÉ IMPLEMENTATION (CPPV INDIVIDUALS METADATA “SAMPLE 1”)	45
FIGURE 18 PROTÉGÉ IMPLEMENTATION (CPPV INDIVIDUALS METADATA “SAMPLE 2”)	46
FIGURE 19 PROTÉGÉ IMPLEMENTATION (CPPV INDIVIDUALS METADATA “SAMPLE 3”)	46
FIGURE 20 PROTÉGÉ IMPLEMENTATION (CPPV CLASS INDIVIDUALS “SAMPLE 1”)......	47
FIGURE 21 PROTÉGÉ IMPLEMENTATION (CPPV CLASS INDIVIDUALS “SAMPLE 2”)......	47
FIGURE 22 PROTÉGÉ IMPLEMENTATION (CPPV CLASS INDIVIDUALS “SAMPLE 3”)......	48
FIGURE 23 PROTÉGÉ IMPLEMENTATION (CPPV CLASS INDIVIDUALS “SAMPLE 4”)......	48
FIGURE 24 PROTÉGÉ IMPLEMENTATION (CPPV CLASS INDIVIDUALS “SAMPLE 5”)......	49
FIGURE 25 CPPV CKAN IMPLEMENTATION OF PUBLIC POLICIES’ ENTRIES	50
FIGURE 26 CPPV CKAN IMPLEMENTATION OF PUBLIC POLICIES’ ENTRIES (PUBLIC POLICY SAMPLE 1).....	51

FIGURE 27 CPPV CKAN IMPLEMENTATION OF PUBLIC POLICIES' ENTRIES (PUBLIC POLICY SAMPLE 2).....	51
FIGURE 28 CPPV CKAN IMPLEMENTATION OF PUBLIC POLICIES' ENTRIES (PUBLIC POLICY SAMPLE 3).....	52
FIGURE 29 UNSTRUCTURED DATA ANALYSIS (TEXT MINING)	54
FIGURE 30 ORIGINAL DATASET METADATA (CATALOGUE) AND COLLABORATION MINING READY DATA MODELS	54
FIGURE 31 PUBLISHER COLLABORATION NETWORK.....	54
FIGURE 32 TOP NAMED ENTITIES DESCRIBING THE OPEN DATA PORTAL “DATA.GOV.IE”	55
FIGURE 33 TOP NAMED ENTITIES TYPES DESCRIBING THE OPEN DATA PORTAL “DATA.GOV.IE”	55
FIGURE 34 TOP NAMED ENTITIES DESCRIBING DATA POSTED BY TOP PUBLISHERS TO THE OPEN DATA PORTAL “DATA.GOV.IE”	56
FIGURE 35 PUBLISHERS COLLABORATION NETWORK OF OPEN DATA PORTAL “DATA.GOV.IE” WITH RELATION STRENGTH > 20.....	57
FIGURE 36 PUBLISHERS COLLABORATION NETWORK OF OPEN DATA PORTAL “DATA.GOV.IE” – SHOWING HIGHEST RELATION STRENGTH SCORE BETWEEN “MARINE-INSTITUTE” AND “GEOLOGICAL-SURVEY-OF-IRELAND”.....	57
FIGURE 37 PUBLISHERS MINED RELATIONS OF OPEN DATA PORTAL “DATA.GOV.IE”.....	58
EQUATION 2 ACCURACY MEASURES.....	59
TABLE 8 ACCURACY MEASURES RESULTS TABULAR VIEW.....	60
FIGURE 38 ACCURACY MEASURES COMPARED.....	60
FIGURE 39 RUNNING PUBLIC POLICY DATA EXTRACTION AND CLEANSING MODULE “CMD COMMANDS AND OUTPUT”	61
FIGURE 40 PUBLIC POLICY DATA EXTRACTION AND CLEANSING MODULE SAMPLE OUTPUT IN MONGO dB BROWSER “PUBLIC POLICY DATA STRUCTURE IN MONGODB”	61
FIGURE 41 PUBLIC POLICY DATA EXTRACTION AND CLEANSING MODULE SAMPLE OUTPUT IN MONGO dB BROWSER “PUBLIC POLICY AS TEXT”	62
FIGURE 42 PUBLIC POLICY DATA EXTRACTION AND CLEANSING MODULE SAMPLE OUTPUT IN MONGO dB BROWSER “PUBLIC POLICY AS WORD VECTORS”.....	62
FIGURE 43 RUNNING PUBLIC POLICY TOPIC MODELLING MODULE “CMD COMMANDS AND OUTPUT”	63
FIGURE 44 PUBLIC POLICY TOPIC MODELLING MODULE SAMPLE OUTPUT IN MONGO DB BROWSER “PUBLIC POLICY TOPICS VECTORS [2 TOPICS]”	63
FIGURE 45 PUBLIC POLICY TOPIC MODELLING MODULE SAMPLE OUTPUT IN MONGO DB BROWSER “PUBLIC POLICY TOPIC VECTOR”	64
FIGURE 46 RUNNING PUBLIC POLICY CO-OCCURRENCE MATRIX BUILDING MODULE “CMD COMMANDS AND OUTPUT”	64
FIGURE 47 PUBLIC POLICY CO-OCCURRENCE MATRIX BUILDING MODULE SAMPLE OUTPUT IN MONGO DB BROWSER “PUBLIC POLICY CO-OCCURRENCE MATRIX”	65
FIGURE 48 RUNNING PUBLIC POLICY RELATED TWEET COLLECTION MODULE “CMD COMMANDS AND OUTPUT”	66
FIGURE 49 PUBLIC POLICY RELATED TWEET COLLECTION MODULE SAMPLE OUTPUT IN MONGO DB BROWSER “PUBLIC POLICY RELATED TWEETS”	67

FIGURE 50 RUNNING PUBLIC POLICY MATRIX BASED RELATED TWEETS QUERYING MODULE “CMD COMMANDS AND OUTPUT”.....	68
FIGURE 51 PUBLIC POLICY MATRIX BASED RELATED TWEETS QUERYING MODULE SAMPLE OUTPUT IN MONGO DB BROWSER “PUBLIC POLICY Co-OCCURRENCE CELL RELATED TWEETS”.....	68
FIGURE 52 PUBLIC POLICY MATRIX BASED RELATED TWEETS QUERYING MODULE SAMPLE OUTPUT IN MONGO DB BROWSER “PUBLIC POLICY Co-OCCURRENCE CELL RELATED TWEETS” DOCUMENT VIEW.....	69
FIGURE 53 PUBLIC POLICY SENTIMENT ANALYSIS MODULE SAMPLE OUTPUT IN MONGO DB BROWSER.....	69
FIGURE 54 PUBLIC POLICY SENTIMENT ANALYSIS MODULE SAMPLE OUTPUT IN MONGO DB BROWSER “DOCUMENT VIEW”.....	70
EQUATION 3 CITIZEN SATISFACTION ANALYSIS FUNCTIONS	71
FIGURE 55 CORE PUBLIC POLICY VOCABULARY IMPLEMENTATIONS IN PROTÉGÉ AND CKAN.....	73
FIGURE 56 DATASETS SHARED BETWEEN MARINE INSTITUTE AND GEOLOGICAL SURVEY OF IRELAND AROUND THE CONCEPT POLLUTION	75
FIGURE 57 DATASETS SHARED BETWEEN MARINE INSTITUTE AND GEOLOGICAL SURVEY OF IRELAND AROUND THE CONCEPT HYDROGRAPHY	76

TABLE OF TABLES

TABLE 1 ARABIC NATURAL LANGUAGE PROCESSING CHALLENGES	8
TABLE 2 CURRENT STATE-OF-THE-ART IN ARABIC NLP RESOURCES AND TOOLS.....	9
TABLE 3 CPPV CLASSES i.e. CONCEPTS	25
TABLE 4 CPPV OBJECT PROPERTIES i.e. RELATIONS	26
TABLE 5 URI NAMING PATTERNS.....	27
TABLE 6 “SIMPLIFIED” REPRESENTATION OF GOV-DAF KNOWLEDGE BASE IN TURTLE	36
TABLE 7 COMPETENCY QUESTIONS.....	40
TABLE 8 ACCURACY MEASURES RESULTS TABULAR VIEW.....	60

TABLE OF EQUATIONS

EQUATION 1 Gov-DAF CITIZEN SATISFACTION FORMULAS ‘SIMPLIFIED’	3
EQUATION 2 ACCURACY MEASURES.....	59
EQUATION 3 CITIZEN SATISFACTION ANALYSIS FUNCTIONS	71

1 INTRODUCTION

Citizens to Government interactions had significant changes over the last decade since the emergence of web 2.0 and social media. Noticeable governmental efforts are on-going for utilizing the social media for better Citizens-to-Government Interaction to ease the conversion from e-government to we-government. There is an increasing need for harnessing social trends in governmental departments' policies for the purposes of raising citizen satisfaction [17]. Using citizen satisfaction data to predict future acceptance or citizen satisfaction of future policy proposal is potentially very beneficial to the government. However, there is currently no established approach for this. Manually monitoring social web contents is ineffective and can consume considerable human resources. Hence, a system for automatic acquisition of citizen opinions, classification, and aggregation is critical for government decision making.

This research work, which is grounded on Natural Language Processing, Machine Learning, Semantic Web and Statistics, aims to fill this innovation gap. The proposed Public Policy Management Framework will offer dynamic, rich analysis of Public Policies, allowing not only faster and easier decision-making process but also more accurate decisions by facilitating the inclusion of public policy collaboration and citizen satisfaction variables in the decision-making process. While citizen satisfaction is rarely employed in government decision-making even when citizen surveys are available, we argue that decision opinions and perception on related public policy issues provide an important input into the decision process. This is important, given that the success of public policies and public services are measured through citizen satisfaction survey. However, acquiring the required citizen satisfaction data through surveys on an on-going basis is resource intensive and infeasible for the government.

In this regard, employing tools which automatically capture citizen satisfaction and opinions such as NLP and Machine Learning tools make this goal feasible. The use of similar tools is reported in [18] in which an e-Participation classification method employing Associative Networks, Spread Activation, and Unsupervised Learning was described *this Method lacks text analysis, offering citizen satisfaction variable to*

decision maker, and the visualization of the analysis results. . Our goal here is to exploit NLP, ML and Web 0.2 techniques for public policy analysis, and a statistical predictor for predicting citizen opinions and sentiments “Gov-DAF” Figure 1 using the ontological classes and instances defined in our designed knowledge base “CPPV” Figure 2.

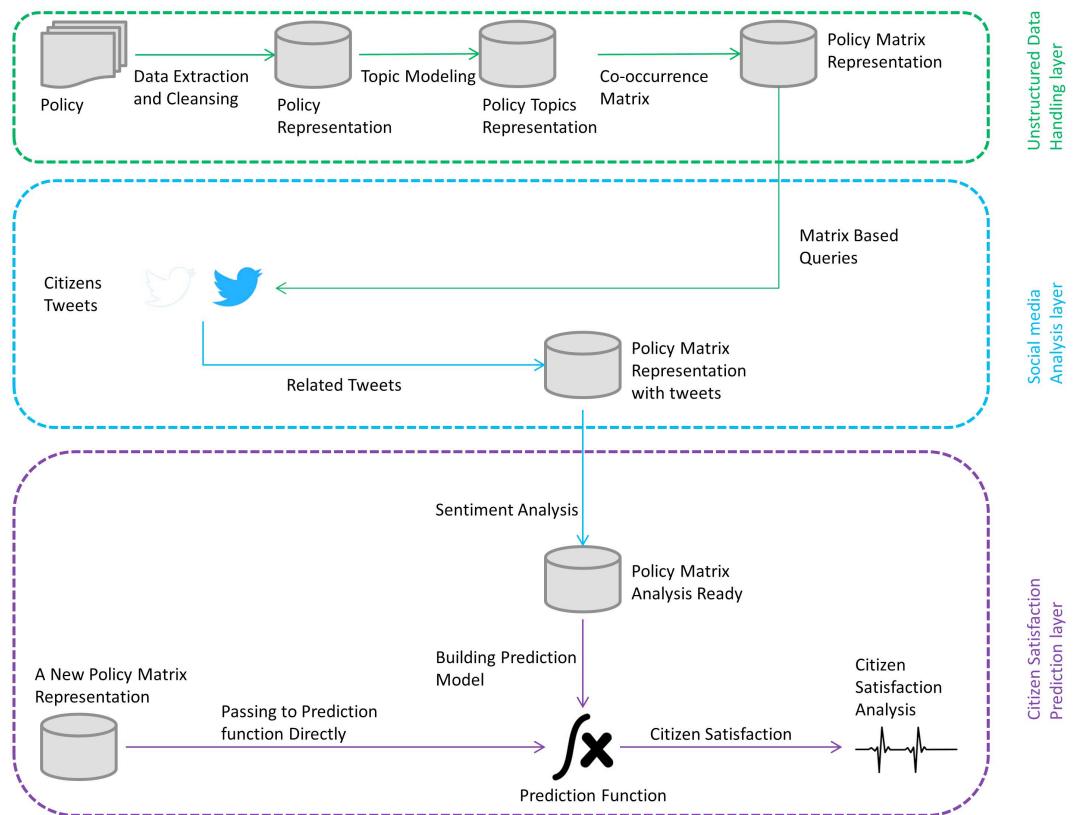


Figure 1 Citizen Satisfaction Analysis Pipeline Gov-DAF

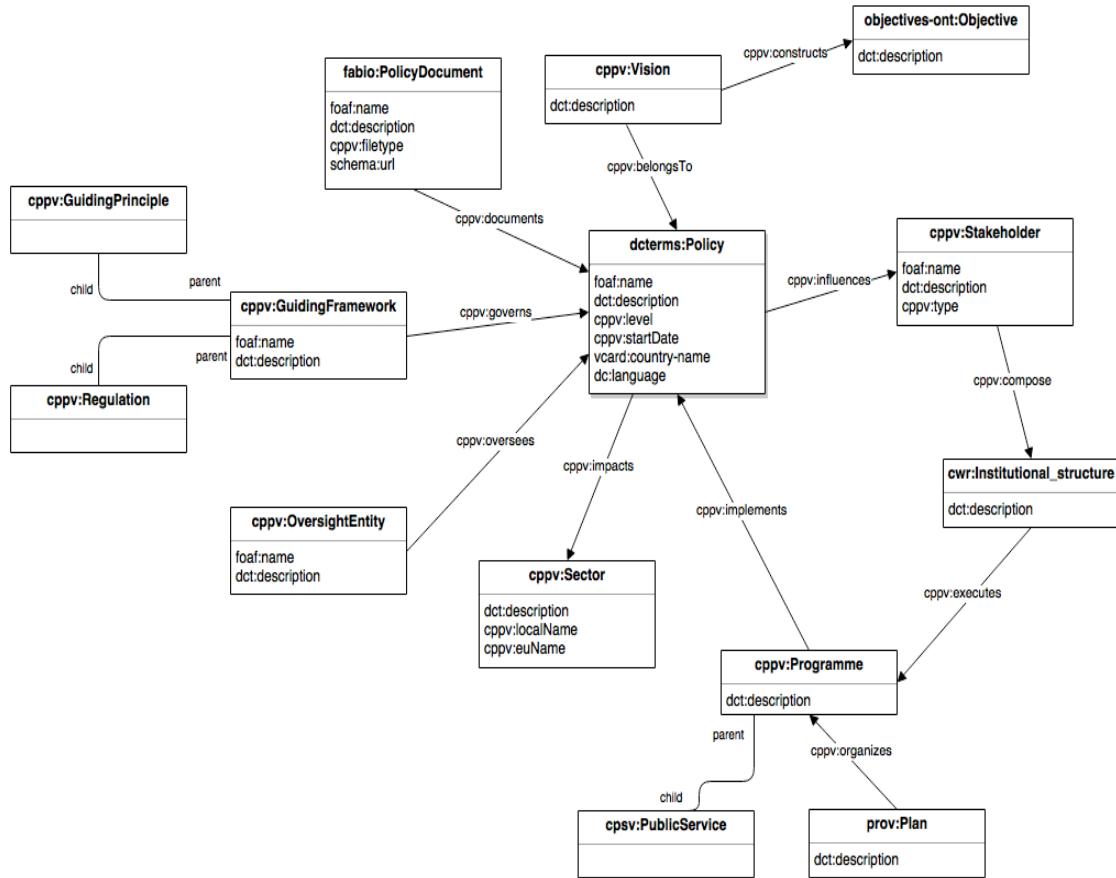


Figure 2 Core Public Policy Vocabulary (CPPV) UML [13]

In 2012 Bandari et al [17] introduced their news article popularity forecasting model, that model predicted the news article popularity variable which is measured by the number of news article url sharing in twitter. They used four independent variables to build the predictive model; those variables are news source, news category, article subjectivity and named entities mentioned in the news article. In our Public Policy Analysis Framework we are building two predictive models for forecasting the actual public policy acceptance rate as dependent variable this variable is quantified using Actual satisfaction rate function Equation 1, Against the independent variable in this case social media public policy acceptance that is quantified using Micropost satisfaction rate function that uses the sentiment analysis of the tweets as the scoring method Equation 1.

To analyze the tweets sentiments to solve the Micropost satisfaction rate equation Gov-DAF is extending multiple algorithms and tools for building a solution pipeline in Figure 1and Figure 3.

Equation 1 Gov-DAF Citizen Satisfaction Formulas ‘simplified’

$$\text{Microposts Satisfaction Rate} = \sum_{e1:en} \frac{\text{positive tweets}}{\text{positive tweets} + \text{negative tweets} + \text{neutral tweets}}$$

$$\text{Actual Satisfaction Rate} = \sum_{e1:en} \frac{\text{positive surveys}}{\text{positive surveys} + \text{negative surveys} + \text{neutral surveys}}$$

Public Policy Ontology Modeling	Computational Analysis of Citizen Opinions and Sentiments	Knowledgebase Construction	Prediction
<ul style="list-style-type: none"> • Investigated Public Policy • Origin and Branch Keyword Extraction. • Policy Aspect Detection. 	<ul style="list-style-type: none"> • Harvesting Citizens Interactions using Twitter. • Opinion Mining. 	<ul style="list-style-type: none"> • Populating our ontology with keywords. • Attaching citizen satisfaction indexes. 	<ul style="list-style-type: none"> • Mining the accumulated knowledge to calculate citizen satisfaction rates towards policy aspects.

Figure 3 Gov-DAF Abstracted [11], [12]

Despite the availability of data loaded into open data portals worldwide¹ [19], [20], methods to maximize public data and policies stakeholders’ engagement and ease data integration and public policy makers collaboration still not complete [21]–[23]. We believe that a proper mining of public policy makers collaboration channels within a single data portal internally as well as between multiple open data portals are not introduced yet. Our work is aiming to develop an open data portals collaboration channels mining framework as shown in Figure 4. To achieve this, we start with data acquisition by harvesting metadata of datasets published on the portal then restructure and store them in MongoDB². Afterwards we construct textual representation from the dataset metadata’s unstructured features, apply DBpedia [24] Named Entity Recognition pipeline called DBpedia Spotlight [25] to extract information that represent those dataset and their publishers i.e. Public policy possible collaborators as well. After that we end up with a semantically enriched dataset upon which we can apply our profiling [23] and collaboration opportunities analysis.

¹ <http://opendatabarometer.org/>

² <https://www.mongodb.com>

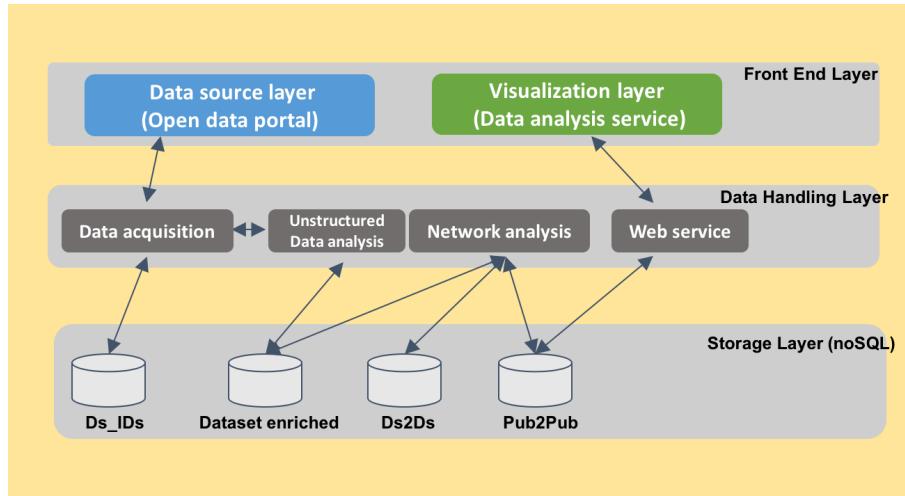


Figure 4 Open Data Catalogues and Publishers' Semantic Profiling Conceptual Framework.

1.1 Research objectives

This research is contributing the following:

- (A) A ready to use Decision Support Systems for Governmental Organizations by using NLP Techniques and Knowledge Management Processes for e-participation and C2G interactions Insights and Analysis
- (B) Filling the following gaps:
 - 1. Government Support System that: A. Allow instant qualitative and quantitative analysis of e-participation and C2G interactions using AI tools. B. Ease and allow the insertion of citizen satisfaction variable into decision equation by presenting analysis results in a visualised dashboard. C. Optimized Public Policy Knowledgebase. D. Public Policy Collaboration Mining Tool
 - 2. Comparison of Arabic Natural Language Processing Analysis methods in our research aspects.

1.1 Research methodology

This research will be carried by applying the Design Science research methodology Figure 5.

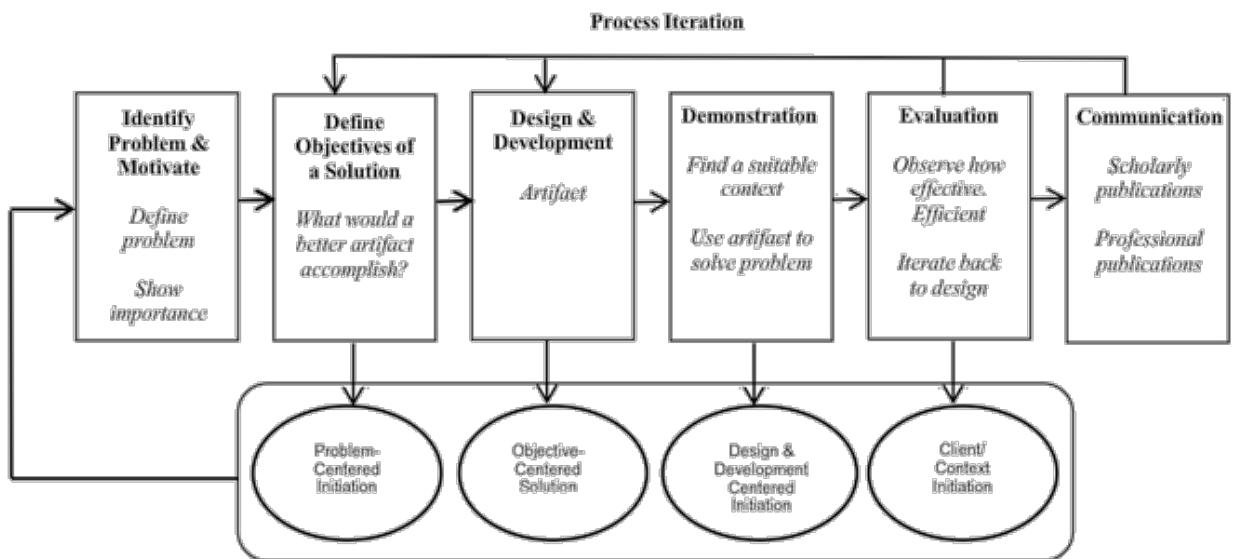


Figure 5 Design Science research methodology

As it both introduces a solution to solve C2G web based interactions analysis automation (Problem Centric: We intend to solve the e-participation and C2G interaction analysis problem), and to Determine the feasibility of using Arabic NLP tools (Objective Centric: We intend to report the accuracy results comparison between Arabic translation and Arabic owl as Arabic text analysis methods).

1.2 Scope and limitations

The scopes of this research are C2G web based interactions mainly social media contents for Gov-DAF component, Open Data Portals contents for Public Policy collaboration Mining component and Public Policies content for CPPV component.

The limitations of the research are mainly the tight coupling with the accuracy and advance of the NLP, ML, Web 0.2 and statistical modules used across the framework components. More discussion on limitation will be found in “Conclusion” chapter.

1.3 Organisation of the dissertation

This Masters Thesis flows on the following structure: Chapter 1: “Introduction”, Chapter 2: “Literature Review, Chapter 3: “Solution Components”, Chapter 4: “Experimentation and Evaluation” and Chapter 5: “Conclusion”.

2 LITRATURE REVIEW

This literature review comes in the following points, in addition to the literature reviews included in the Solution Components, and Experimentation and Evaluation chapters.

2.1 Background

Decision Support Systems are a variety of tools that assist humans for making decisions in different domains “ Almost anything could qualify as a decision support system - from a sophisticated interactive decision-modelling system, through a programmable hand-held calculator to a cup of coffee.” [26].

Atrial to narrow the definition of Decision Support Systems had been made by Sprague and Carlson in early 80s “a class of information system that draws on transaction processing systems and interacts with the other parts of the overall information system to support the decision-making activities of managers and other knowledge workers in organization.” [27] .

A broader and Abstracted definition by [26], would fit better for describing this wide spectrum systems only sharing the ability to assist human in decision making “1. "Decision" emphasizes the primary focus on decision-making in a problem situation rather than the subordinate activities of simple information retrieval, processing or reporting. 2. "Support" clarifies the computer's role in aiding rather than replacing the decision-maker. 3. "System" highlights the integrated nature of the overall approach, suggesting the wider context of user, machine and decision environment.” [26].

In 2008 Power, Daniel J had published a useful Decision Support Systems History overview [28], presented the previous four decades work on the DSS area.

Power, Daniel J had presented in this research a seven types of DSS were prepared by Steven Alter in his MIT phd back in 1980, this types are as follows “Alter concluded from his research [29] that decision support systems could be categorized in terms of the generic operations that can be performed by such systems. These generic operations extend along a single dimension, ranging from extremely data-oriented to

extremely model-oriented. Alter conducted a field study of 56 DSS that he categorized into seven distinct types of DSS. His seven types include:

- **File drawer systems** that provide access to data items.
- **Data analysis systems** that support the manipulation of data by computerized tools tailored to a specific task and setting or by more general tools and operators.
- **Analysis information systems** that provide access to a series of decision-oriented databases and small models.
- **Accounting and financial models** that calculate the consequences of possible actions.
- **Representational models** that estimate the consequences of actions on the basis of simulation models.
- **Optimization models** that provide guidelines for action by generating an optimal solution consistent with a series of constraints.
- **Suggestion models** that perform the logical processing leading to a specific suggested decision for a fairly structured or well-understood task.” [28].

According to this categorization of Decision Support Systems, this research will fit under Data Analysis Systems.

2.2 Knowledge Management Process

This research applied Knowledge Management Process Phases: Knowledge Creation, Knowledge Storage and Retrieval, Knowledge Transfer and Knowledge Application as reviewed and introduced in [30] to produce knowledge from Citizen to government interactions, in a disciplined manner.

2.3 Artificial Intelligence Tools

We applied different Artificial Intelligence tools in different parts of our research. Natural Language Processing (NLP) techniques introduced and developed in many literatures for text analysis. Our research utilizing the following NLP research and more: (A) Named Entity Recognition tools e.g. Stanford Named Entity Recognizer [31] and Alchemy Entity Extractor reviewed in [32]. (B) Sentiment Analysis or opinion mining tools e.g. Sentiwordnet [33] and WEKA assisted [34] or a combined approach [35]. (C) Knowledge Bases and ontologies to assist in text analysis e.g.

Dbpedia [24]. (D) Semantic Similarity tools to be used in citizen to government interactions classification and categorization into governmental departments and topics e.g. extracting DIstributionally related words using Co-occurrences (DISCO) introduced in [36], [37].

More tools and algorithms are studied, deployed and reported in Solution Components, and Experimentation and Evaluation chapters.

2.4 A Comparative Study of Different Arabic Language Analysis Methodologies

Arabic is the one of the widest spoken language in the world, with over 200 million speakers, utilized by twenty four countries [38]. In our research we looked intn how to serve Arabic language speaking countries with our public policy management framework. Our initial approach was to fit best available Arabic Nataural Language Processing tools into different pipeline modules wthin our solution components. Following a discussion of the challnengs that Arabic Natural Language Processing tools facing and Current state-of-the-art in Arabic NLP resources and tools. Followed by a discussion of why and why not to fit those tools into our public policy management framework.

2.4.1 Challenges of Arabic NLP

Following a list of Arabic NLP challenges and it's related works that is discussing the challenges and/or proposing solutions for it.

Table 1 Arabic Natural Language Processing Challenges

<i>Challenges</i>	<i>Related Works / Surveys</i>
Highly Ambiguous Arabic Words	[39]
Absence of Rigorous Standards of Written Text	[39]
Corpus	[40]–[44]
Arabic diacritization	[45]
Arabic dialects	[42], [46]
Machine Translation	[47]

2.4.2 Current state-of-the-art in Arabic NLP Resources and Tools

Following a list of related works discussing Current state-of-the-art in Arabic NLP Resources and Tools related to our public policy management pipeline .

Table 2 Current state-of-the-art in Arabic NLP Resources and Tools

NLP Module	Arabic NLLP Related Works / Surveys
Sentiment Analysis / Opinion Mining	[40], [41], [48]–[50]
Text Summarization	[51], [52]
Named Entity Recognition (NER)	[52]–[55]
Part of Speech Tagging	[56]
Stemming	[57]

2.4.3 Fitting Arabic Natural Language Processing into our Public Policy Management Framework

Following challenges we faced to fit Arabic NLP into our Framework:

- Work related to Arabic NER is rather limited and under-studied [53].
- Arabic Sentiment Analysis approaches have continued to evolve despite lack in Arabic linguistic resources [40].
- In Modern Standard Arabic texts are typically written without diacritical markings. The diacritics are important to clarify the sense and meaning of words. Lack of these markings may lead to ambiguity even for the natives [45].
- Many Arabic applications, such as machine translation, text-to-speech, and information retrieval, are vulnerable due to lack of diacritics [45].
- Part of Speech tagging is a main component of most NLP modules and in Arabic Language a few researches have been investigated in Arabic part of speech tagging with modest effort [56]
- The Arabic language is one of the Semitic languages; it exhibits a very systematic but complex morphological structure based on root-pattern schemes [57].

- The complexity of the automatic identification of Arabic dialects since. No clear border can be found between the dialects, but a gradual transition between them [46].
- We also considered using Arabic machine translation tools to generate English contents and pass it to the designed pipeline but there is no machine learning technique that fully meets human requirements. Finding a quick and efficient translation mechanism [47] to use was never successful.

Finally, The Arabic-specific challenges are mainly caused by Arabic morphological complexity, limited resources and dialects, while the general linguistic issues include polarity fuzziness, polarity strength, implicit sentiment, sarcasm, spam, review quality and domain dependence [50]. We found No Arabic Language based tool for Topic Modeling or Opinion Mininig using Deep Learninig which are main parts of our public policy management framework.

2.5 Citizen to Government Relation Analysis

This research is concerned with the analysis of Citizen to Government interactions. Which is related to e-Participation, e-Participation is defined as: "eParticipation is defined here as the participation of individuals and legal entities and groups thereof in the decision-making process in the branches of government using information and communication technology (ICT) equipment. In the context of the federal government's eGovernment activities, we interpret this along two lines, i.e. • as an offer to participate, in a form upgraded by the use of ICT, in plans and decisions by the federal ministries and their downstream public agencies and by the German parliament, • as an extension of applicable rules and requirements for other levels of government within the federal government's (framework) legislative powers. In international debate, a distinction is made between the term "eParticipation" and the term "eDemocracy" in that the latter also covers elections as the most binding form of citizen participation. Therefore, this study does not deal with eVoting." [58]. And also defined as: "Therefore, our working definition of eParticipation is: 'eParticipation describes efforts to broaden and deepen political participation by enabling citizens to connect with one another and with their elected representatives and governments by using Information and Communication Technologies (ICT).' " [59]. By Citizen to Government we describe the relation conducted or the interactions between citizens

and government through Social Media (SM) as reviewed in [60]. Those concepts' importance for governmental decisions effectiveness is proved by the quantity of research efforts spent to maximize their benefits to government performance among others [61]–[65], and the huge efforts spent for shaping eParticipation research domain done by insight centre researchers among others [16], [66]–[72], [15]. Previous research for analyzing e-Participation and Citizen to Government interactions [18], introduced a classification method for classifying e-Participation and Citizen to Government interactions to certain domains using, Associative Networks, Spreading Activation and Unsupervised Learning. This lack the analysis of the text, offering citizen satisfaction variable to decision maker, and the visualization of the analysis results.

2.6 Text analysis accuracy measuring methodology

NLP researches evaluate their work by comparing analysis results to human analysis of the same data set by crowdsourcing, and define their accuracy based on the distance between matching judgments with humans and mismatching judgment with human. This method have a disciplined approach, starting with measuring the accuracy of the human judgment it self before running the test. By running a pre-evaluated test and measure human accuracy based on it. This accuracy percentage will be used in all proceeding accuracy tests made by this human [73].

2.7 Conclusion

In this chapter we partially presented main related literature areas. A more specialized and comprehensive literature reviews are reported in Solution Components and Experimentation and Evaluation chapters according to their related component of our research. In the following chapter “Solution Components ” we elaborate more on our proposed solution components for Public Policy Management challenges and issues.

3 SOLUTION COMPONENTS

The focus of this chapter will be the Gov-DAF components proposed to fulfil the research questions. Gov-DAF is proposing a public policy management framework. Which is inclosing three different models and tools addressing public policy management issues. First we answered the public policy modelling issue by designing, authoring, and implementing the Core Public Policy Vocabulary (CPPV) [16], [15], [14]. Then we answered the public policy collaboration issue by authoring the Semantic Profiling Collaboration Mining Tool [66]. Finally we proposed a citizen satisfaction measuring tool to estimate current as well as in progress public policies success index [11], [12], [14]. Following a detailed explanation of the Gov-DAF components listed above:

3.1 Core Public Policy Vocabulary

Public policies possess a vital role in multiple levels of a nation's life, those policies are shaped by a nation's tenets, ideologies and principles, and shaping a nation's notoriety, lifestyle and public satisfaction. Public policies learning are a compulsory element of Government to Government learning field, furthermore public policies analysis represent a research objective for both pure scientific research and business research. Thence, in this paper we propose a framework that will abate the commotion that public policy researchers face during the policy finding and analyzing processes. Our core public policy vocabulary CPPV, were conceived upon a purposeful manual policy analysis and aiming to profound semantically enabled and federated policy repository to harness the public policy analysis difficulties. Specifically via utilizing CPPV through population, a wide spectrum of use cases can be realized e.g. Semantic Policy Search Engine, Government Learning and Predictive Policy Analytics.

3.1.1 Component Overview

Nationals and government officials raise large requests on the improvement concerning e-government. There are various, occasionally conflicting, hopes and objectives towards increased strategy responsiveness, transparency, efficiency,

legitimate security, coordination, interoperability, protection also unrestricted data circulation. Those improvements of e-government encounter a differentiated assessment matrix. This transform the various egovernment application improvement missions into a puzzling task including strategies and public policies. Thus comprehensive information and skills are necessities to fulfill improvements requests. E-government studies are able to support the perception and improvement about e-government complex ecosystem.

A significant number of Scrutinize literature generating important assistances to tackle a variety of challenges and difficulties. There are public policy researches about domestic and global public programs and plans. Also legitimate studies of potentials and difficulties in egovernment progress have been done.

There are diverse egovernment actions and systems assessments. The varied hopes and objectives stated above have formed research from different angles; e.g. standardization and integration angle. From such angle research like Core Public Service and Core Public Policy have emerge.

Public policy is an important method for observing citizen and administration interaction. Hence the public policy research is very important and fundamental for the design and implementation of egovernment ecosystems.

This paper describes the methodology, implementation and results of designing the core public policy vocabulary and accommodating ontology. The ontology is expected to have the following conclusions:

Demonstrating that the core public policy vocabulary can be used as a foundational RDF (Resource Description Framework) vocabulary to standardize public policy data that originates from local, regional, national and international e-government portals.

Demonstrating that the definition of uniform HTTP (Hypertext Transfer Protocol) URI (Uniform Resource Identifier) sets for public policies facilitates information management. Proving that the use of standardized controlled vocabularies can significantly facilitate the search for public policies, especially in a multilingual context.

3.1.2 Component Background and Related Work

3.1.3 Public Policy

A "policy" is composed of an administration's action or a group of actions, then the government "make", "implement" or "carry out" the policy same as doing a decisions. Similar to decisions the policy is not an announcement, nor only a group of actions, still, as with decisions, we are able to conclude what an individual's or entity's policy is either using the declaration he announces concerning the policy, or, even if no announcements have been made or we don't trust his announcements due to his contradicting actions. However, similarly, if the announcement or the group of actions are deceptive people are able to say that this is not compliant with the "true" policy [74].

A "policy" is the situation where a domestic, provincial or nationwide governmental administration makes an announcement of commitment to accomplish a single objective or set of objectives. A policy might be detailed using laws or other formal documents to discuss the institutional settings e.g. executing organization, and rewards/fines system. [75]

Public policy is usually inferred as the aggregation of actions, legal acts, governing measures, duties, sequences of operations and resource allocation precedence that are completed, in the 'public's' name [76].

In the mid-1960s the developing of public policy was to be a leading subfield under the domain of political science. Generally, the investigation of public policy started with the civilization itself. Public policy is the study of administration's choices and activities intended to react to an issue of the community. Public policy examinations are meant to assist administration by providing them with valuable insights through public policy deep study [77].

3.1.4 Public Policies Types

According to [77] , Public Policy types are classified according to administration's control approach.

Patronage, regulatory, and redistributive policies are the control types in which an administration's control approach is exercised.

“Patronage policies: incorporate those administration activities that give motivators to people or organizations to attempt exercises they would just reluctantly attempt without the guarantee of a reward. These promotional techniques can be classified into three types: subsidies, contracts, and licenses.

Regulatory policies: allow the government to exert control over the conduct of certain activities. The most obvious examples of regulation techniques include civil and criminal penalties for certain behaviours.

Redistributive policies: control people by managing the economy as a whole. The techniques of control involve fiscal (tax) and monetary (supply of money) policies.

Fiscal techniques: use tax rates and government spending to affect total or aggregate demand. Each particular approach to taxing or spending can have a different impact on the overall economy, so political entrepreneurs often propose or initiate policies with the goal of achieving specific impacts.

Monetary techniques: used by the Federal Reserve Board (the “Fed”), also try to regulate the economy by changing the rate of growth of the money supply or manipulating interest rates.”

3.1.5 Public Policies Job

Among government tasks public policy development holds a high importance rank. Public policy development is essential in order to build a plan to achieve one or more of the following governmental objectives:

Increasing the country’s wealth, dropping unemployment rates, sustaining economic growth and public incomes i.e. the economic policy.

Decreasing inflation or motivating economic growth through manipulation of public spending and/or public income i.e. the fiscal policy.

Providing good health care system, equality, welfare, education opportunities to the public i.e. the social policy.

Governing international relations (foreign policy) [75].

3.1.6 Vocabularies

Vocabularies contribute to the semantic web framework by enabling data integration e.g. when having multiple data sets using different naming patterns for

features, or to use vocabulary added data to conclude a fact that is not provided in the original data. Considering the use of ontologies in the health care domain. Health specialists use them to model information about symptoms, diseases, and treatments. Pharmacological corporations use vocabularies to model knowledge about drugs, dosages, and allergies. Merging the above knowledge from the Health communities with patient information allows a wide variety of smart applications e.g. decision support systems recommending possible treatments. [78]

3.1.7 Ontology

Ontology in Philosophy means the study of the nature of being, but ontology in Artificial Intelligence have a different meaning, defined as a formal explicit specification of a shared conceptualization [79]. An ontology is modelling a domain, by detecting types of entities in the domain and relation types connecting them, also it shall be accepted and standardized by domain specialist [80], [81].

Five tasks of ontologies are to perform the following: organizing and structuring information, reasoning and problem solving, semantic indexing and search, semantic integration and interoperation and understanding of the domain [82].

[83] defined the ontology as “Ontology is an explicit and formal shared abstract view of a part of the real world. This view is described by a whole of tools as a vocabulary formed of concepts, relations, axioms and rules of inference” Accordingly, the ontology founds a standard vocabulary for the domain specialists for generating, merging, exchanging, accessing and reusing knowledge.

3.1.8 Related Work

First of all we need to clearly distinguish between “Formal Policies” and “Public Policies”, the need and the importance of Policy Formalization are illuminated by the following examples on Formal Policies in Health Information Systems, and Distributed Systems domains.

In the United States EMR (Electronic Medical Record) and HIS (Health Information System) are governed by a complex number of laws, regulations and policies from the federal, state and local governments as well as institutional policies and business agreements. Guaranteeing steady harmonization of policies across federal, state, and institutional levels remains a significant challenge [84]. In addition

to the complex nature of the issue, the policies and regulations are revised and altered from time to time. Increasingly stringent enforcement requirements with increasing penalties for nonfulfillment add to the substantial challenge for the development of health information systems. To overcome these issues researchers have been concentrating on formalization of policies that control health information systems.

Another example of Formal Policy in Distributed Systems domain is the idea of implementing machine-enforceable policies as a way of specifying and governing the behaviour of distributed systems emerged as a result of increasing system and network management complexity [85]. By means of changing policies, a system can be continuously attuned to accommodate variations in externally imposed limitations and environmental conditions. Through these applications, policies have been exploited either to automate access control or to manage network administration tasks, such as configuration, security, recovery, or quality-of-service (QoS). New application fields for policy management include, among others, multi-agent systems, pervasive and mobile devices, and autonomic computing systems [86]–[91].

Lastly, Formal policies are relying on the same originating principles as Public Policies, but having a dissimilar context of usage and implementation. This should finalize and clarify the difference between the two types of Policies.

The need and the importance of the modelling of Law and Public Policy have been shown since the post-war decades. Public Policies are increasingly demanded as a preferred tool to promote collective action. Nowadays, these public policies are developed through sophisticated participatory schemes comprising of a variety of actors, public or private. Indeed in the present context of globalization, though the State and its administration are crucial stakeholders [92], their impact fades gradually into a more diffuse institutional environment [93] involving a multitude of other stakeholders [94]. At both ends of the spectrum of governance for example, we discover on one hand the growing involvement of supranational entities and on the opposite, the involvement of nongovernmental organizations. Following the same trend it becomes compulsory to base projects of public policy on impact analyses themselves[95], [96] regulated by legal provisions e.g. [97]. The modelling of law and public policy is an evolving field of research targeting in particular at understanding their "functioning", at simulating *ex ante* their possible impacts, at finding ways of realizing a wanted goal (back casting), at creating tools for decision support, at

enabling the participative drafting of legal texts or policy, particularly by using ICT (Information and Communication Technology) or methods developed for the analysis and structuring of social networks.

A proof of the current research gap and the potential contribution of CPPV to support and enhance the need for public participation in policy-making is presented. According to the OECD ‘all OECD member countries distinguish new ICTs to be prominent tools for increasing citizen engagement in public policy-making’ since ‘the unparalleled degree of interactivity offered by new ICTs has the probability of the expanding the scope, breadth and depth of government consultation with citizens and other key stakeholders in the course of policy making’ [98]. While the literature recognizes and emphasizes the great potential of ICT for assisting and enhancing public participation in policy-making, it has not been adequately investigated in more detail so far. For example, in [99] the OECD expresses the absence of knowledges to date with the legislation practice as the public policy go through to be permitted, should reach improved public engagement through sophisticated ICT support. The major research question is to discover the best means of using argumentation systems and ontologies for increasing public engagement.

Existing examples of policy modelling: There are many researches and solutions targeting policy modelling among others [100]–[104], and the following are examples of these work.

Education policy enactment: While this example is exclusively intended for the Educational domain, it still offers a deep understanding for the policy modeling process and methodology. In recent times, the concept of context has been deployed for education policy research focused on enactment, rather than to policy implementation [105]. In this case, [106] use the term “enactment” to refer to the dual processes of policy interpretation and translation by a diverse array of policy actors in a wide variety of situations and practices. The first process, interpretation, indicates a preliminary reading and making sense of, or meaning of policy texts. The second process, translation, [105] propose a re-reading of policy, literally ‘enacting’ policy, in and through talk, school plans, meetings, classroom lessons, data walls, school websites. The deployment of context as an analytic device for understanding the processes of policy enactment enables researchers to address the complex means in which official

policies are enacted within and across schools [107]. Through the conceptual lens of context, an emphasis on theorizing enactment is thus an appropriate corrective and important contribution to the critical policy studies literature. What [105] target to achieve in their analytic use of context to comprehend policy enactment is ‘a grounded interpretation of the diverse variables and factors (the what), as well as the dynamics of context (the how) that shape policy enactments and thus to relate together and theories interpretative, material and contextual dimensions of the policy process’. They try to offer a framework that can take into account ‘... a set of objective conditions in relation to a set of subjective “interpretational” dynamics’ and suggest four overlapping and interrelated ‘contextual dimensions’. In addition, [105] propose that contexts are always specific, dynamic, and can shift in and outside schools. Their conceptual toolkit surrounding policy enactment and context shapes on previous analytic concepts, mainly notions of performativity to analyze the policy cycles of the education reform package(s) designed to re-professionalize teachers, and redefine what encompasses good teaching, effective schooling and quality learning [108].

In the OCOPOMO (Open Collaboration Policy Modelling) project ³, they are targeting better engagement of stakeholders in policy making using a novel approach for engaging stakeholders in policy development is conceptualized and implemented [109]. Public policies are investigated and modelled such as renewable energy policy of the Kosice region in Slovakia, housing policy of the city of London or the distribution of structural funds in the Campania region in Italy. In the overall OCOPOMO ontological structures, the CCD aims to make modelling decisions comprehensible and traceable for stakeholders. In this work the authors have presented the CCD meta-model, which is the fundamental vocabulary and relation for describing policy contexts. They additionally sketched the concept for semi-automatic transformation of conceptual policy descriptions of the CCD into formal policy models.

A dichotomy regarding policymaking arises usually and globally, as a result of the fact that design and implementation are two quasi-independent and lightly linked subprocesses. Policy analysis and design constitutes a top-down process comprising of

³www.ocopomo.eu/

political -or highly politicized- personnel. It is outcome oriented and functions on the basis of general criteria such as mission and vision concepts, organizational and environmental values and strategies, political priorities etc. Alternatively, policy implementation is mainly a bottom-up process involving P.A. professionals: middle management and low-level personnel. Therefore, implementation processes are directed by intra-organizational priorities and day-to-day management requirements and boundaries. They are short-term output oriented with only ambiguous references to the “big picture” and loose links to the organization’s long-term goals and strategic priorities and goals.

The unfitting connection and mismatching of two complementary steps of a process thought to be linear, leads to the disintegration of the general rational policymaking framework and produces incoherent, ineffective and inefficient practical policies. Public policies become even less effective insofar as the vertical gap couples with a horizontal one, as is frequently the case. Policy outcomes are typically the conjunction of individual processes in diverse policy fields. The failure of full communication and cooperation amongst all these distinctive policy constituencies and networks more often than not consequences in poor policy outcomes. Loose and ineffective horizontal communications is a very common and extensive defect of contemporary policymaking and administrative systems [110].

The challenging situation referred to above creates an crucial need for a bi-dimensional integrative interface linking both the vertical and the horizontal policy gaps and connecting all policy field actors to a cooperating “policy community”. This type of interface can be constructed through extensive use of modern informational and communication technologies applied at each and every phase of the general policy making process.

One side of this process must be associated with the transcription of existing policy making procedures in ICT applications, so as to homogenize, simplify and accelerate vertical coordination and horizontal networking and, consequently, facilitate policy integration. Generic process and data structures, a high level model for strategic planning to afford common definitions, vocabulary and conceptual framework for policy making within wider policy field-based sub- categories of public organizations and a unifying “enterprise architecture” on which all public administration processing

and information systems ought to be based in order to become interoperable are some vital requirements for integrated public policies.

[111] Focuses his work on the internal policy decisions being made for the formation of cultural policies within distinct cultural organizations. Despite the fact that the impact of these on the policies that are being targeted by a range of other, external, policy actors from central to local governments, to external funding agencies, to a variety of community and voluntary organizations are deemed important in their own right, they are not the central topic of attention. Therefore, precisely why government departments made the individual decisions that they did regarding topics varying from cultural regeneration strategies, culture as a tool for social inclusion, or the elevation of cultural excellence is not necessarily the central issue: instead, the attention is on how staff inside the organizations that are affected by these decisions cope with them in the context of their own decision and policy-making strategies. This discussion and analysis is enclosed within the context of the overall argument within the social sciences on the subject of the interplay between structure and agency.

A collective argument within this context looks at the relative significance of structure and agency for the functioning of individual and organizational decision processes, with policy actors being either, like Macbeth, ‘cabined, cribbed, confined, bound in’ by a set of relatively invariant structural limitations, or adept of claiming that ‘we are not a government poodle’ [112], able, instead, to make their own independent decisions through the exercise of political agency. The degree to which such positions are either tenable or relevant to an understanding of structure and agency in the context of social action somewhat depends upon the underlying ontological, epistemological, theoretical and methodological positions that analysts assume [113]. In the situation where a realist ontological position is assumed, where agency and structure are perceived as continuously entangled [114]–[117], with each being acknowledged as central to an effective analysis of policy activity, then the relative significance of conflicting structural and agential factors in any given distinct case can only be determined empirically and is unchangeable to definitional *fiat*. Of necessity such empirical analysis must include both structure and agency to be effective, which brings about questions regarding the appropriateness of differing methodology for making sense of their interplay.

3.1.9 Methodology

Conceptual modelling (CM) is the elicitation and the representation of the general knowledge that an information system operating in a specific domain needs to know [118], [119]. Describing a domain of the real world through conceptual models means viewing it in a particular way, i.e. through the assumption that the world consists of concepts, e.g. entities, objects, events, processes [118], [120]. In particular, conceptual modelling aims at representing static (e.g., objects, entities) and dynamic phenomena (e.g., events and processes) in a particular domain [121]. CM is an essential part of information systems development [118], [122]–[124] which traditionally consists of the analysis, design, and implementation stages Figure 6. Information systems analysis transforms a perceived real-world domain or universe of discourse [125] into a conceptual model, while design and implementation transforms the model, into a design model and an information system eventually [122].

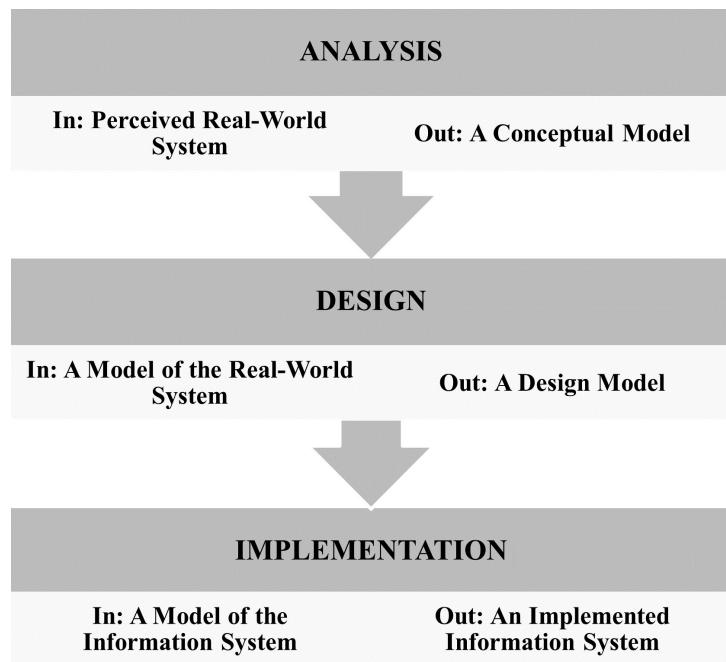


Figure 6 The Role of Conceptual Model in Systems Development [122]

CM encompasses four elements [121]:

CM grammar [122] (or CM notation [126]) that provides a set of constructs and rules showing how to combine the constructs in order to model the real world domain [121]. The IS literature includes a vast variety of conceptual modelling

grammars including data oriented, process oriented and object-oriented ones, [118], [121], [123], [125]–[127].

CM method (or modelling technique [126], [128] that prescribes how to map the real phenomena of the domain into a model of the domain [121]. A rather limited number of CM methods exist in the literature [121], [123].

CM script (or conceptual schema [118], [119], [123], [124], [126] or conceptual model [123], [126] is the output of the CM method and is constructed using a CM grammar [121]. It is usually one (or more) diagrams [123] depicting the domain's concepts. A great variety of conceptual scripts have been proposed in the literature [84], [118], [121], [123], [128].

CM context, which is the setting in which conceptual modelling occurs and scripts are used.

To study the vocabulary design process and outline gaps and opportunities we followed a Systematic Literature Review methodology. SLR is an important piece of any theoretical investigation as it generates a strong basis for the research [129]. The research modules implemented is represented in Figure 7.



Figure 7 Systematic Literature Review [130]

Afterwards, we followed a vocabulary design methodology shown in Figure 8 adapted from the Core Public Service Vocabulary methodology [131].

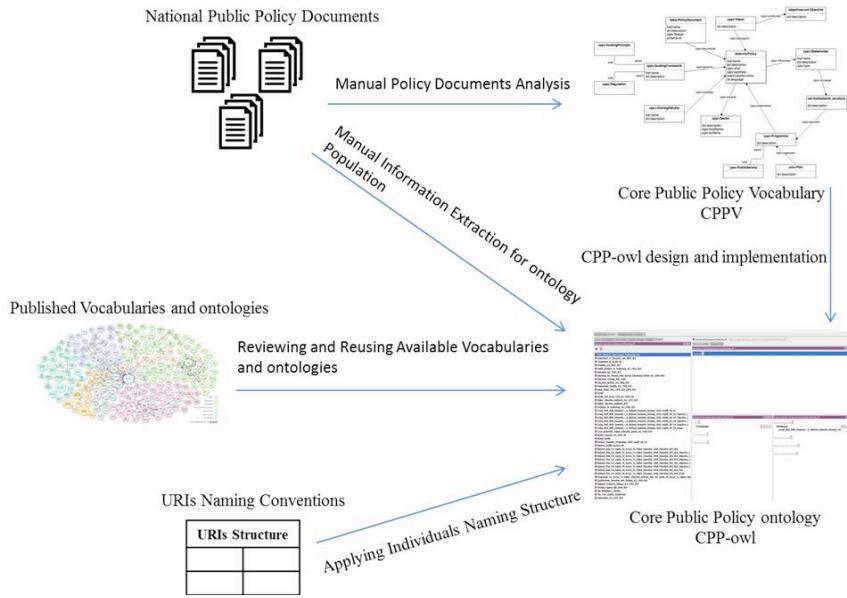


Figure 8 The Core Public Policy Vocabulary Modelling Methodology

The work started by assessing general use cases and a number of public policy documents that illustrate the action plan for some European countries. These have been used as primary sources of requirements collection to envisage the definition of the core elements of the public policies catalogue. (as shown in Figure 8), a successive fine tuning of the model has been undertaken in order to reflect the actual characteristics of real life public policies available from a variety of administrations (bottom-up approach). In particular, we considered public policies offered by administrations, and governments. We also took into account the list of policies envisaged in the experimental phase that is currently activated for the development of the Core Public Policy Vocabulary. This bottom up approach led to changes of the preliminary data model and to the definition of a second version of it that we called Core Public Policy Vocabulary. When defining classes and properties, classifications have been identified; e.g., the main theme of a policy, the channels through which policy can be provided to end-users, the types of input documents, etc. Common and shared classifications are key elements for enabling semantic interoperability and linking among different types of data [132]. To this end, we started to define controlled vocabularies using already available classifications. This approach allows us to provide harmonization between policy and data upon which policies can be built.

3.1.10 Vocabulary Design

As this research aiming to design and implement a cross domains lightweight ontology that can describe and imitate a real life public policy. We presented the Core Public Policy Vocabulary (CPPV), and the Core Public Policy ontology (CPP-owl). The following lines will discuss both specifications and evaluation of our vocabulary, and the Core Public Policy ontology (CPP-owl).

3.1.11 Conceptual Model

Our Core Public Policy Vocabulary (CPPV) contains the following classes and properties to imitate the real life public policy. We concluded these elements through intensive manual investigation of cross-domains public policies. Afterwards we manually populated the Core Public Policy Ontology (CPP-owl)⁴ ontology by extracting and mapping information from real life public policies as shown in Figure 2. A detailed description of the CPPV elements is listed in Table 3, Table 4 and Table 5.

Table 3 CPPV Classes i.e. Concepts

<i>Class</i>	<i>Description</i>	<i>URI</i>
cppv:GuidingFramework	The Parent Class for the Guiding Principle and the Regulation.	http://www.egov.deri.ie/cppv.owl#GuidingFramework
cppv:GuidingPrinciple	The foundation of every organization. They define what is truly important for its success.	http://www.egov.deri.ie/cppv.owl#GuidingPrinciple
cwr:institutional_structure	Explicit and implicit institutional rules and policies designed to provide a structure where various work roles and responsibilities are delegated and coordinated. source:(cwr)	http://www.fao.org/aims/aos/cwr.owl#institutional_structur e
objectives-ont:Objectives	Things policy actions set out or are intended to attain or accomplish. Top level class for Goals and Objectives. source:(objectives-ont)	http://www.daml.org/experiment/ontology/objectives-ont#Objectives
cppv:OversightEntity	A distinct or independent body charged with supervision and observation of policy implementation.	http://www.egov.deri.ie/cppv.owl#OversightEntity
terms2:PeriodOfTime	An interval of time that is named or defined by its start and end dates. source:(dublin core)	http://dublincore.org/documents/2012/06/14/dcmi-terms/?v=ter ms#PeriodOfTime
prov:Plan	A detailed proposal for doing or achieving	http://www.w3.org/ns/prov#P

⁴ <http://vmogi01.deri.ie/egovcppv/>

	something. A plan is an entity that represents a set of actions or steps intended by one or more agents to achieve some goals. source:(prov)	lan
cppv:Policy	A combination of law, rules and guidelines that determine or control results, management and information usage – that designs the role of information in a community.	http://www.egov.deri.ie/cppv.owl#Policy
fabio:PolicyDocument	A document embodying a policy that describes and defines how something should be done. source:(fabio)	http://purl.org/spar/fabio/PolyDocument
cppv:Programme	A planned series of future events or performances. Structures for implementing and executing policies.	http://www.egov.deri.ie/cppv.owl#Programme
cpsv:PublicService	This class represents the service itself. A public service is the capacity to carry out a procedure and exists whether it is used or not. It is a set of deeds and acts performed by or on behalf of a public agency for the benefit of a citizen, a business or another public agency. source:(cpsv)	http://purl.org/vocab/cpsv#PublicService
cppv:Regulation	The system of rules which a particular country or community recognizes as regulating the actions of its members and which it may enforce by the imposition of penalties.	http://www.egov.deri.ie/cppv.owl#Regulation
cppv:Sector	A distinct part, especially of society or of a nation's economy that is impacts or impacted by the public policy.	http://www.egov.deri.ie/cppv.owl#Sector
cppv:Stakeholder	All key parties influenced or affected by the implementation of a policy.	http://www.egov.deri.ie/cppv.owl#Stakeholder
cppv:Vision	An organisation's road map, indicating both what it wants to become and guiding transformational initiatives by setting a defined direction for growth or achievement.	http://www.egov.deri.ie/cppv.owl#Vision

Table 4 CPPV Object Properties i.e. Relations

<i>Object Property</i>	<i>Domain</i>	<i>Range</i>	<i>Is inverse of</i>
belongsTo	Vision	Policy	hasVision
compose	Stakeholder	institutional_structure	composedOf
constructs	Vision	Objective	constructedBy
documents	PolicyDocument	Policy	documentedby

executes	institutional_structure	Programme	executedBy
governs	GuidingFramework	Policy	governedBy
impacts	Policy	Sector	isImpactedBy
implements	Programme	Policy	implementedBy
influences	Policy	Stakeholder	influencedBy
organizes	Plan	Programme	organizedBy
oversees	OversightEntity	Policy	overseenBy
relatedTo	Thing	Thing	relatedTo

3.1.12 URI Naming Patterns

Regarding the URI building guidance introduced in [133], [134] we designed a URI naming patterns to assure a representative and reusable URIs are generated through ontology accommodation phases.

Table 5 URI Naming Patterns

Object of	URI Naming Pattern
Policy	Error! Hyperlink reference not valid.
Programme	Error! Hyperlink reference not valid.
Sector	<u>http://{domain}/{Sector'sEUName}</u>
Stakeholder	Error! Hyperlink reference not valid.}
OversightEntity	Error! Hyperlink reference not valid.}
Plan	<u>http://{domain}/Plan/{PolicyName}{PublicationYear}_{Sector'sEUName}_{CountryName}_{PrimaryOrganisation}</u>
Regulation	Error! Hyperlink reference not valid.
PublicServices	Error! Hyperlink reference not valid. Adopted from CPSV [133]
GuidingPrinciples	<u>http://{domain}/GuidingPrinciples/{PolicyName}{PublicationYear}_{Sector'sEUName}_{CountryName}_{PrimaryOrganisation}</u>
Vision	Error! Hyperlink reference not valid.
Objectives	Error! Hyperlink reference not valid.
PolicyDocument	Error! Hyperlink reference not valid.
InstitutionStructure	Error! Hyperlink reference not valid.

3.2 Public Policy Collaboration Mining for Citizen Satisfaction

To increase policy acceptance we developed a new policy collaboration tool that will enable a well analysed and developed public policies according to the increased governmental collaboration. Due to the increasing adoption of open data among governments worldwide especially in the European Union area, a deeper analysis of the newly published data is becoming a mandate. Apart from analyzing the published dataset itself we aimed on analyzing published dataset catalogues. A dataset catalogue or a dataset metadata contains features that describe what the data is about in a textual representation. So, we first acquire data from open data portals, choose descriptive dataset catalogue features, and then construct an aggregated textual representation of the datasets. Afterwards we enrich those textual representations using Natural Language Processing (NLP) methods to create a new comparable data feature “Named Entities”. By mining the new data feature we are able to produce datasets and publishers relatedness network. Those networks are used to point similarities between the published data across multiple open data portals. Pointing all possible collaborations for integrating and standardizing data features and types would increase the value of data and ease its analysis process.

3.2.1 Component Overview

Despite the availability of data loaded into open data portals worldwide⁵ [19], [20], methods to maximize stakeholders’ engagement and ease data integration still not complete [21]–[23]. We believe that a proper mining of collaboration channels within a single data portal internally as well as between multiple open data portals are not introduced yet. Our work is aiming to develop an open data portals collaboration channels mining framework as shown in Figure 4. To achieve this, we start with data acquisition by harvesting metadata of datasets published on the portal then restructure

⁵ <http://opendatabarometer.org/>

and store them in MongoDB⁶. Afterwards we construct textual representation from the dataset metadata's unstructured features, apply DBpedia [24] Named Entity Recognition pipeline called DBpedia Spotlight [25] to extract information that represent those dataset and their publishers as well. After that we end up with a semantically enriched dataset upon which we can apply our profiling [23] and collaboration opportunities analysis.

3.2.2 Component Background and Related Work

Following concepts definitions and a literature review of correlated research areas Open Government Data, NLP and Collaboration Mining:

3.2.3 Open Government Data

Open Government Data referred to the datasets generated and published by governmental departments “without any restrictions on its usage or distribution” and it doesn't contain any personal or undisclosed data [135]. OGD vary by multiple aspects for example: a) OGD publishing department or agency domain e.g. Agriculture Data, Transport Data, Environmental Data, Financial Data and Telecommunication Data. b) Data format e.g. Excel, Text, PDF, CSV, Theoretically, Government Open Data is operational or administrative governmental data available to use, redistribute, and analyze “in any form without any copyright restrictions” [136]. Regarding the open government working group draft in 2007⁷ they generated initial open data principles: data must be complete, primary, timely, accessible, machine-processable, nondiscriminatory, nonproprietary, and license-free. Then they generated further open data principles, data must be online and free, permanent, trusted, assumed to be open, documented, safe to open, and designed with public input Figure 11 shows the Irish government's open data portal which we used for our experiments⁸.

⁶ <https://www.mongodb.com>

⁷ http://public.resource.org/8_principles.html

⁸ <http://data.gov.ie>

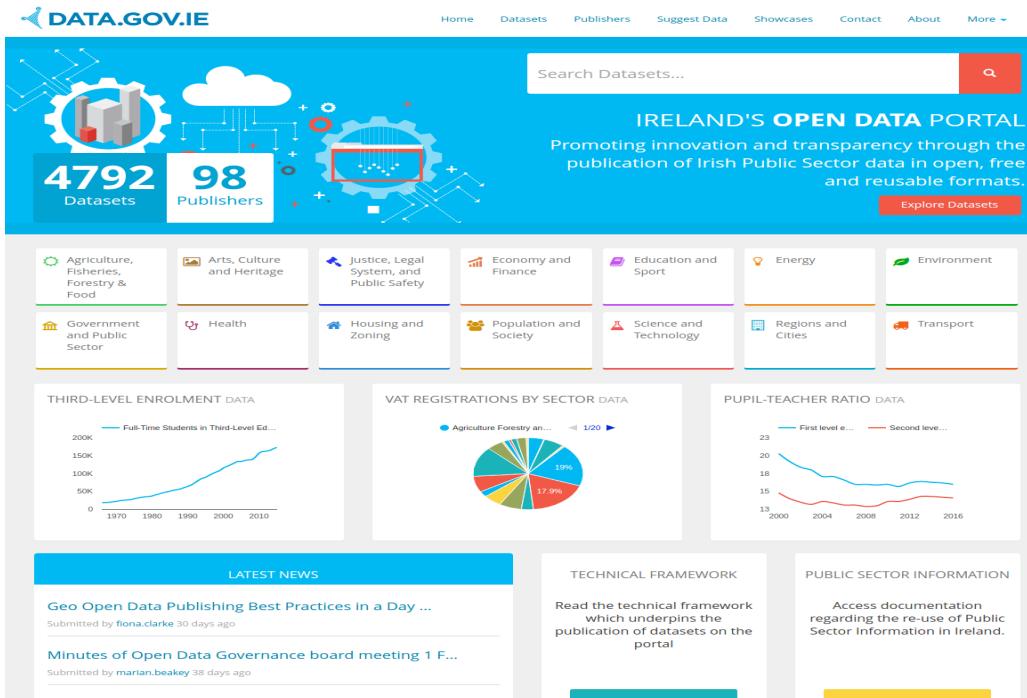


Figure 9 Irish Government’s Open Data Portal

3.2.4 Natural Language Processing

Following we discuss the correlated features of Natural Language Processing to our research. Specifically, Named Entity Recognition applications:

3.2.5 Named Entity Recognition

Named Entity Recognition is the process of discovering Named Entities (NE) laying within a given text, a common definition of NE is as follows [55], “an information unit described by the name of a person or an organization, a location, a brand, a product, a numeric expression including time, date, money and percent found in a sentence.”[137]. NER applications are implemented using multiple methodologies:

The Supervised Learning techniques use a big manually categorized dataset. Then this dataset is used for training the recognition algorithm. Supervised Learning techniques apply Conditional Random Fields [138], Hidden Markov Models [139], Decision Trees [140], Support Vector Machines [141] and Maximum Entropy Models [142]. The objective of these methods is to identify and categorize related key-words. The

unavailability of manually categorized datasets and the high cost of generating them, represent a challenging obstacle against Supervised Learning Techniques.

The Semi-Supervised Learning and Unsupervised Learning techniques use either a small categorized dataset for training the algorithm [143], or a clustering based algorithm. Further Unsupervised Learning techniques depend on lingual resources e.g. WordNet, and statistics to solve the NER task as a prediction problem [144].

3.2.6 Natural Language Processing in E-Government

There are few implementations of NLP technologies in the e-government area. Examples from the works found: A proposed application for gathering crime data from police departments and eyewitness stories and apply NLP technologies with GATE [145]. A system that imitate email answering process automatically or semi-automatically using NLP technologies [146]. Another application presents an original model for incorporating multimedia data to assist e-government tasks [147].

3.2.7 Mining for Collaboration

In general, due to the great benefits and possibilities of collaboration opportunities mining and discovery research e.g. Process speed enhancing, Standardization and Integration. The detection of possible collaboration opportunities within an organization or across multiple organizations and platforms is targeted in multiple domains. Following the few existing work digging into mining for collaboration area: Mining for collaboration in library domain, the research is harnessing the detection of possible collaboration opportunities with academic professional based on their publications to increase the benefits of students [148]. Collaboration mining between governmental levels and departments based on their objectives, resources and services to increase the government efficiency regarding public policy development and implementation, crisis management, etc. [149]. Collaboration mining tool using agent technology to analyze the collaboration between information on the web to help the tool users to get their desired materials more accurately and faster [150]. Collaboration mining of team members using summaries of successful past projects to increase moderator efficiency to promote project partner's awareness of best way to formulate a proposal for a European research project [151].

3.3 GOV DAF

Citizens' satisfaction index anent public policies is a core political research question. Seeking agile and efficient public policies, canny public policy makers are perpetually investigating how to weigh citizens' satisfaction index towards their policies, in order to overhaul faulty public policy aspects or topics that outcome a negative citizens' satisfaction index. Multiple automated and manual citizen's satisfaction index calculation methods are applied [1]–[10], bearing in mind that a correct calculation will grant a significant success for the public policy and the public policymakers. Our perspective is that nevertheless the index was well calculated or not it is always too late, public policy is already issued and citizen reactions are ensue. Hence, we previously proposed a public policy satisfaction prediction framework [11], [12], this framework reckon on a knowledge base that allow formulating the prediction formulas. To acquire this knowledge base we extend the Core Public Policy Vocabulary [13], apply Named Entity Recognition and Topic Modelling in parallel for keywords extraction and semantic similarity measuring for relating the detected keywords with the pre-defined public policy aspects. The aforementioned methods will allow an automated population of the prediction knowledge base.

3.3.1 Component Overview

In 2012 Bandari et al [17] introduced their news article popularity forecasting model, that model predicted the news article popularity variable which is measured by the number of news article url sharing in twitter. They used four independent variables to build the predictive model; those variables are news source, news category, article subjectivity and named entities mentioned in the news article. In Gov-DAF “ A Government Decision Analytics Framework Based on Citizen Opinion” we are building two predictive models for forecasting the actual public policy acceptance rate as dependent variable this variable is quantified using Actual satisfaction rate function in Equation 1, Against the independent variable in this case social media public policy acceptance that is quantified using Micropost satisfaction rate function that uses the sentiment analysis of the tweets as the scoring method in Equation 1. To analyze the tweets sentiments to solve the Micropost satisfaction rate equation Gov-DAF is

extending multiple algorithms and tools for building a solution pipeline in Figure 1 and Figure 3 .

Gov-DAF knowledge base pipeline implementation adopted the following vocabulary and methodologies: (1) CPPV. (2) Initial approach “Named Entity Recognition Based Methodology” of building the Gov-DAF knowledge base pipeline as introduced in [11], [12]. (3) Enhanced approach “Topic Modelling Based Methodology”. Finally we evaluated both methodologies against each other and presented the results. Through both methodologies the following vocabularies, algorithms and tools are extended and/or adopted as detailed below:

Structured Public Policy Indexing using CPPV [13] and CKAN [152]

Gov-DAF reckon on public policies as assets for achieving the ultimate target of obtaining meaningful insights about public policies that help public policymakers. Thus, Semantically structured collection and Indexing of public policies or assets for Gov-DAF analytical purposes was one of the main motivation of creating CPPV [13] as part of the research. CPPV offers semantic indexing of public policies’ (assets) metadata Figure 2, and CKAN “ Comprehensive Knowledge Archive Network” [152] the world’s leading open-source data portal platform offers public policy physical documents indexing. Within Gov-DAF Knowledge base building pipeline we extended CPPV with the public policy analytics classes (cppv-ext:AnalyticalAspect, :Keyword), and properties (cppv-ext:type, :occurrence_count, :extends, :composed_of) as the Gov-DAF knowledge base elements that will enable opinion harvesting and analysis in later phases Figure 10.

Public Policy Text Analysis by Named Entity Recognition using Stanford NER [31] and Topic Modeling using LDA [153]

To populate Gov-DAF knowledge base with keywords extracted from public policies we applied two text analysis methods and measured their accuracies indices in our usage domain, which is not necessary indicate their overall accuracy in other usage domains. First method applied for public policy text analysis is Named Entity Recognition using stanford NER [31] where Gov-DAF with the help of stanford NER extracts persons, places and organizations that are composing the public policy. Second method applied for public policy text analysis is Topic Modeling using Mallet implementation of LDA “Latent Dirichlet allocation” [153] here Gov-DAF apply Mallet LDA to cluster keywords composing the public policy into topics vector and

then we apply NER to discover keywords types with the possibility to apply Stanford Entity Resolution Framework [154].

Semantic Relatedness using DISCO [36], [37]

Semantic relatedness is used in two cases, First use case is when NER Methodology is applied as Gov-DAF uses Semantic Relatedness using DISCO “Extracting DIstributionally related words using CO-occurrences” for generating branch keywords Figure 11. Second use case is just before applying opinion harvesting, sentiment analysis and satisfaction estimation according to Gov-DAF Figure 2, Gov-DAF first relate keywords back to certain public policy aspects for deeper public policy analysis in other words for a multi-level satisfaction estimation. Gov-DAF applies Semantic Relatedness using DISCO “Extracting DIstributionally related words using CO-occurrences” for the target of relating extracted public policy keywords with public policy aspects using semantic relatedness based algorithm illustrated in [11], [12].

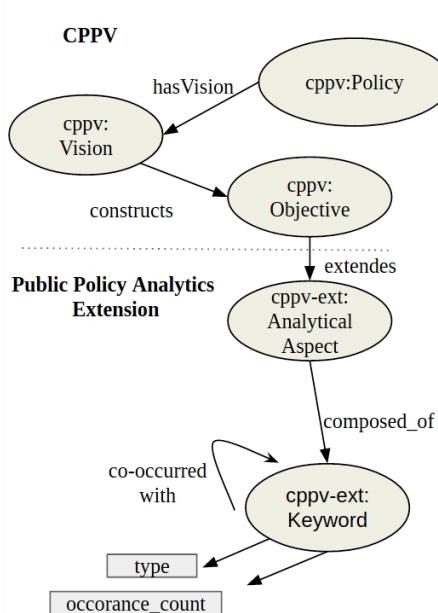


Figure 10 CPPV Policy Analytics Extension

3.3.2 Gov-DAF knowledge base building pipeline

Gov-DAF follows the initial approach “Named Entity Recognition Based Methodology” [11], [12], and also the new enhanced approach “Topic Modelling Based Methodology” for designing Gov-DAF knowledge base building pipeline.

Named Entity Recognition based Methodology was the initial methodology that we presented previously which uses Stanford NER for public policy text Figure 11.

Topic Modelling based Methodology alter the initial Gov-DAF knowledge base building pipeline in public policy text analysis phases, as it uses Mallet LDA for topic modelling to recognize and cluster keywords Figure 12.

Gov-DAF knowledge base building pipeline has two implementations as discussed above, following are a close-up of the Gov-DAF knowledge base building pipeline components:

3.3.3 Approach independent Gov-DAF knowledge base building pipeline stages

Public Policy input

Gov-DAF inputs or assets are the public policies required to be analyzed along with the public policy analytical aspects i.e. public policy objectives entered by the domain experts. A policy can be either an old public policy that is under analysis or a new public policy that is under discussion to be introduced, public policies documents can be inputted in many document formats e.g. pdf.

Inputted public policy is divided into sentences to suit Mallet LDA analysis according to the Topic Modelling Methodology. Public Policy aspects reflect the main components of a public policy. Aspects are defined and fed into our system by domain experts and decided by the user during public policy input phase. A public policy row text, sentences, and aspects vector are the output of this phase.

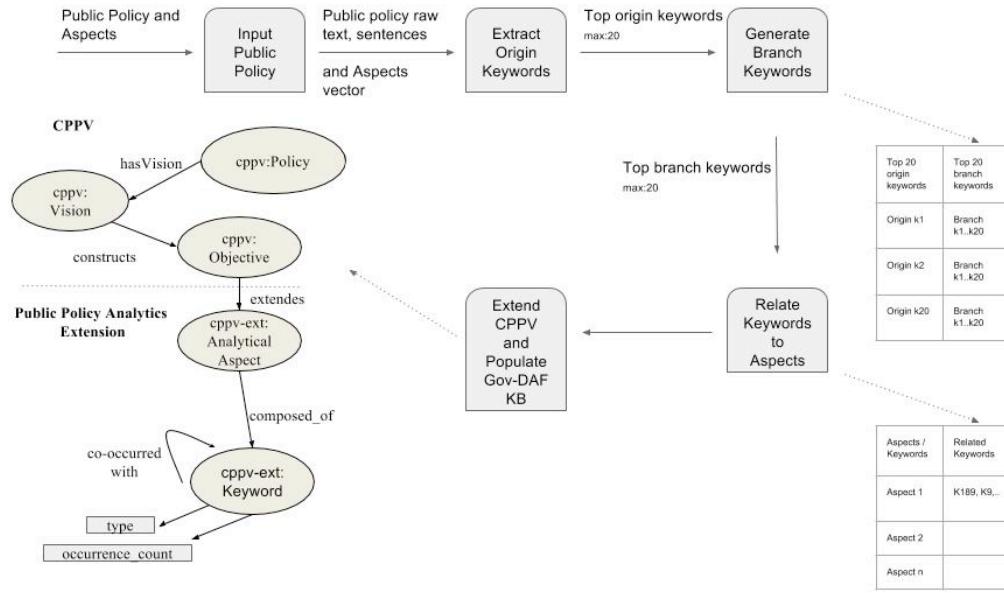


Figure 11 Gov-DAF Knowledge Base Building Pipeline using Named Entity Recognition Based Methodology

Relate Public Policy Keywords to Aspects

Public Policy aspects are to be connected to a set of origin and branch keywords or topic cluster; this set should be strongly descriptive towards the aspect. Keywords will be used for both citizen opinions collection and analysis. Automating this process is carried out using semantic similarity score approach, by calculating the semantic relatedness score between public policy aspects and public policy keywords then nominating top related keywords for every aspect this process quality is measured and reported.

Extend CPPV and Populate Gov-DAF Knowledge Base

Extend the CPPV with public policy analytics extension (cppv-extended) with Class (cppv-extended:Keyword): the keywords extracted from the public policy either using NER method or Topic Modeling method, and properties (cppv-extended:type, occurrence_count) as shown in Figure 2 and Table 6,. Then populate the Gov-DAF knowledge base with both public policy aspects i.e. objectives defined by domain experts and Gov-DAF users

Table 6 “Simplified” representation of Gov-DAF Knowledge Base in Turtle

cppv and cppv-ext	Sample KB entries
cppv:Policy	ex:BioStrategy a cppv:Policy ;
cppv:Vision	ex:vision1 a cppv:Vision;

cppv and cppv-ext	Sample KB entries
	ex:policy1 cppv:hasVision ex:vison1;
cppv:Objective	ex:O1 a cppv:Objective; ex:vision1 cppv:constructs ex:o1;
βcppv-ext:AnalyticalAspect	ex:AA1 a cppv-ext:AnalyticalAspect; ex:O1 cppv-ext:extends ex:AA1;
cppv-ext:Keyword	ex:Education a cppv-ext:Keyword; ex:AA1 cppv-ext:composed_of ex:Education;
cppv-ext:type	ex:Education cppv-ext:type "Person";
cppv-ext:occurrence_count	ex:Education cppv-ext:occurrence_count "100";

Named Entity Recognition Methodology dependent pipeline stages Figure 11

Extract Origin Keywords

Public policy text contains places, persons and organizations within its sentences, all those elements or entities are possible candidates for being the main public policy actors. Using Stanford NER, we recognize entities and tag it with its type for possible multidimensional correlation analysis. Filtering redundant occurrences of those keywords hold their occurrence score to use it as a significance weight of the keyword, sort keywords based on the significance weight and finally use top 20 keywords as our possible main policy actors candidates.

Generate Branch Keywords

After origin keywords recognition and filtration process, we start a keywords network exploration process using DISCO library to extract distributional related words using co-occurrences, DISCO is founded over the similar words clustering algorithm introduced in [155]. Relaxing and expanding the keywords network by nominating top 20 related branch keywords to every origin keywords according to their semantic relatedness score calculated by DISCO library

Topic Modeling Methodology dependent pipeline stages Figure 12

Detect Topics Clusters

Using Mallet LDA topic modelling methodology [153] over public policy sentences extracted in phase one, will allow enhanced approach for Public policy text analysis and keywords extractions and clustering.

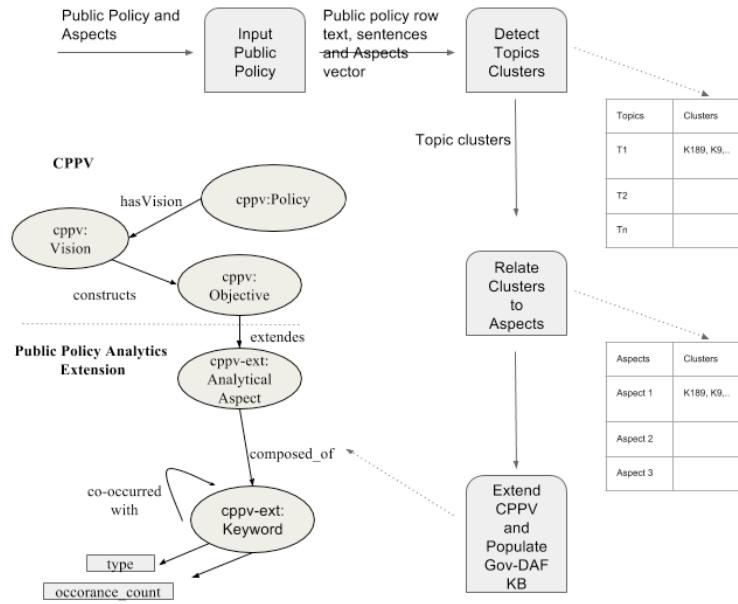


Figure 12 Gov-DAF Knowledge Base Building Pipeline using Topic Modelling Based Methodology

3.4 Conclusion

This chapter “Solution Components” overviewed the design of the proposed Gov-DAF components to fulfil the research questions. Next chapter “Experimentation and Evaluation” will demonstrate the implementation of the Gov-DAF components.

4 EXPERIMENTATION AND EVALUATION

Following Experimentation and Evaluation details of the Gov-DAF components:

4.1 Core Public Policy Vocabulary

4.1.1 Validation

Our research evaluation is based on the following arguments, which are fully satisfied and approved during our work:

Expertise Adoption: During the deferent stages of the research we communicated with policy makers and policy researchers, we presented the research idea and concept and they approved the need for such tool to improve, facilitate and federate the policy making, implementing and acceptance rating processes.

Ontology Validation: Through the stages of conceptual design and ontology implementation we adopted the best practice literature to obtain internal consistency and reusability of our ontology. Then we validated our ontology structure using a set of w3c recommended tools, now we are satisfied of the ontology as it passed the mentioned validation tests.

Competency Questions: After populating the cpp-owl according to the described criteria we designed a set of competency questions according to the literature review and queried our ontology against this CQ and we get 100% accurate answers as shown in Table 7.

CPP-owl passed the validation process that was done using OWL Validator [156] and RDF Validator [157] which are W3C recommended ontology validation tools [158]. Also it passed the internal consistency checks done over Classes, Object proprieties and Data properties. Finally, after applying the competency questions the performance of the CPP-owl for answering the question was 100% accurate.

4.1.2 Competency Questions

Our competency questions built according to the work of [159], [160]

Table 7 Competency Questions

Sample Competency Questions	Correct?
Who are the Stakeholders of { Policy x}?	✓
What policy impacts {Stakeholder x}?	✓
Who oversees {Policy x}?	✓
What are the related Programs to {Program x}?	✓
What Policies start/end in {2017}?	✓
What is the guiding framework of {Policy x}?	✓
What are the Stackholders influenced by {Program x}?	✓
What are the Sectors impacted by {Program x}?	✓

4.1.3 Applications

4.1.4 Abstract Use Cases

Follows the possible use case of the Core Public Policy Vocabulary (CPPV), and Core Public Policy ontology (CPP-owl), also depicted in Figure 13:

As a foundation layer for a Centralized European Semantic Public Policies Repository (CESPPR):

This will be achieved through deploying the proposed Core Public Policy Vocabulary (CPPV), and the Core Public Policy ontology (CPP-owl) on centralized repository and apply a manualized data entry process for capturing and storing different European Public Policies according to the proposed Core Public Policy Vocabulary (CPPV) and the Core Public Policy ontology (CPP-owl). CPPV and CPP-owl usage cases can be divided into two layers; First layer is the technical implementation of the CPPV and CPP-owl at this stage we will have a Centralized European Semantic Public Policies Repository (CESPPR) that allow a wide spectrum of usage cases for many actors with interest of Public Policies. Second layer is the different Categories of usage cases and utilizations of the CPPV and CPP-owl such as Government to Government Learning, Research Projects, and Commercialization.

Semantic Policy Search Engine: At this level a Semantic Search engine will be very tempting research idea, building Public Policy Semantic Search engine (PPSSE) over the resulted semantic repository will enable the emergence of the concept Policy Boutique. By offering more advanced, smart and sophisticated search queries will result a relevant, fined tuned and customized search results. Consequently, Policy Designer can use the Semantic Search Engine (PPSSE) to easily targets and find his needs from the Public Policy Repository as if it was a Public Policy Boutique.

Policy Learning: A Federated European Semantic Public Policies Repository to enable Government to Government learning. The Federation aim of the (CPPV) and the (CPP-owl) will be obtained by offering discovery and access features to the semantically stored European Policies. At the end of this level a Government to Government (G2G) policy learning framework will be enabled to participant and nonparticipant Governments upon agreement wishing to learn or enhance their policy designation skills, procedures or stages. Now a Pan European Cross Border Public Policy Compatibility and Interoperability Infrastructure (PCPPI) became a near target we can achieve with less hassle, due to the Establishment of the following Interoperability Fundamentals: (CPPV), (CPP-owl) and the Centralized and Federated European Semantic Public Policy Repository. At this stage we will be successfully achieving not only a standardized Public Policy storage schema but also a standardized Public Policy Designation Elements. As the Proposed Core and Extension Vocabulary are to be the construction backbone of a standard Public Policy.

Predictive Policy Analytics: At this stage we can apply the Predictive Policy Acceptance Algorithm [161], this will allow a predictive evaluation of the new policy produced using first and second usage cases and may also be used as an evaluation process of our proposed vocabulary.

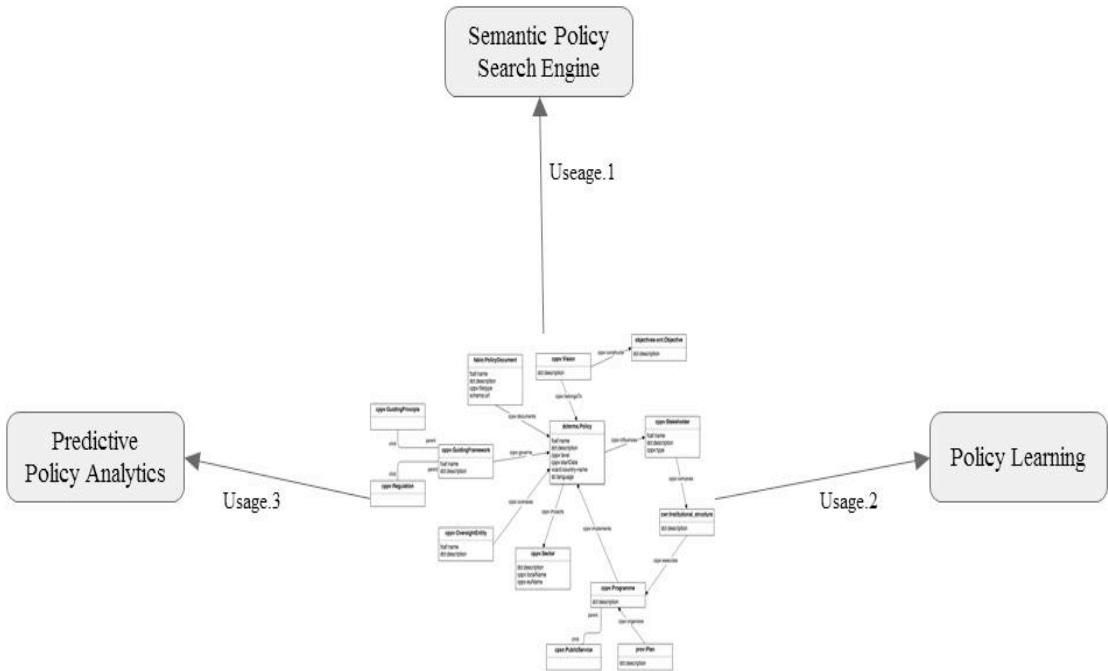


Figure 13 Core Public Policy Vocabulary (CPPV) Use Cases

4.1.5 Scenarios

As discussed in previously we are targeting an extendable set of use cases for our proposed core public policy vocabulary and ontology, for clarifying the idea we expect the following scenarios to be real life scenarios after our research being adopted and used by the European governments.

Scenario 1: "Semantic Policy Search Engine" Researchers and Business users of our ontology may use this search engine to browse the public policies or run complex queries, to discover opportunities in the market weakness in the system etc. Example: Firm X is specialized in the energy sector; they are extending their business to Country Y, to get a better understanding of the market they need to review Country Y's energy policies and its related policies as well to do that they may use our search engine answer their question about this new market.

Scenario 2: "Government Learning" During the policy making or policy revising processes governments may seek help from our ontology to find best practices or certain policy aspects to adopt. Example: Government Y is making a new Health policy they want to review deferent Health polices implemented in the European union, query for certain aspects of the Health policies to

improve their policy making skills and avoid weaknesses in their new policy which will avail a higher citizens' satisfaction rate in Country Y.

Scenario 3: "Predictive Policy Analytics" After making the policy the government may look for a predictive citizens' satisfaction rate against their new policy to allow an early evaluation of the policy at aspects level [161]. Example: Government Y used our ontology to assist them during making their new Health policy, and now they are eager to get feedback or initial evaluation of this policy. They can use our predicative analysis facilities adopted from [161] to fine tune their policy aspects in case of low citizen's satisfaction predicted rates.

4.1.6 Experiments

CPPV was experimented in both technologies Protégé and CKAN, following experiments results:

4.1.7 Protégé Implementation

CPPV RDF/XML and owl Source code can be found at ⁹ and in appendix A. Screen samples in Figure 14 to Figure 24 demonstrate protégé implementation of CPPV.

4.1.8 Protégé Screen Samples

⁹ <http://vmogi01.deri.ie/egovcppv/>

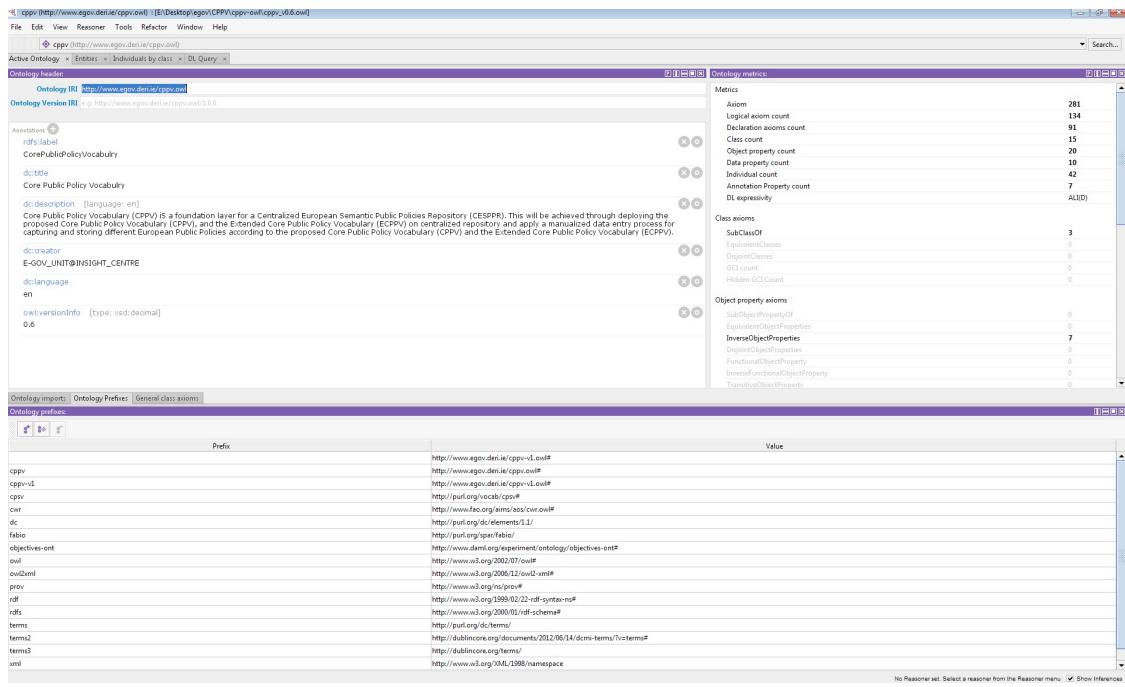


Figure 14 Protégé Implementation (CPPV metadata)

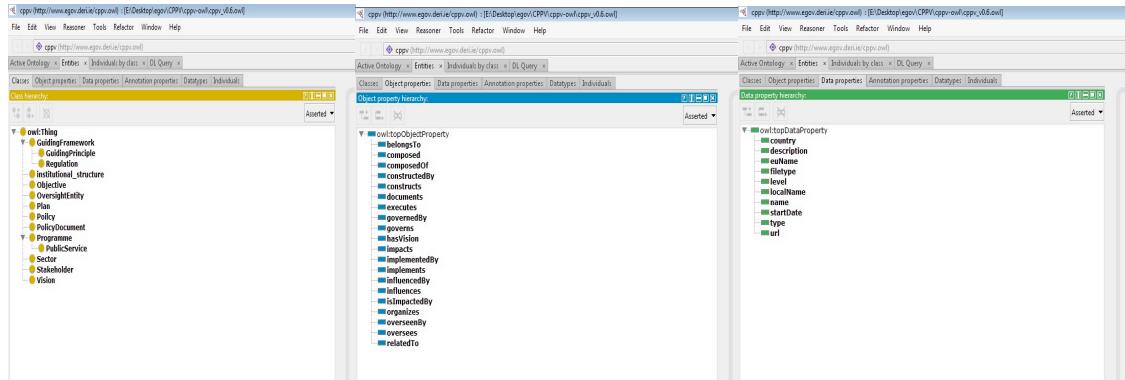


Figure 15 Protégé Implementation (CPPV Classes, Properties and Data types)

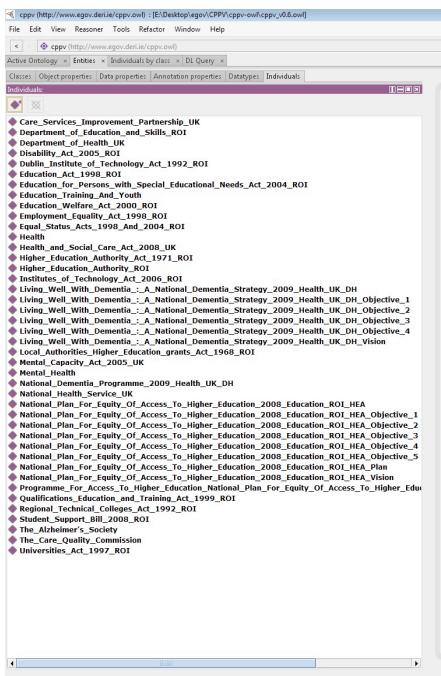


Figure 16 Protégé Implementation (CPPV Individuals)

Figure 17 Protégé Implementation (CPPV Individuals metadata “sample 1”)

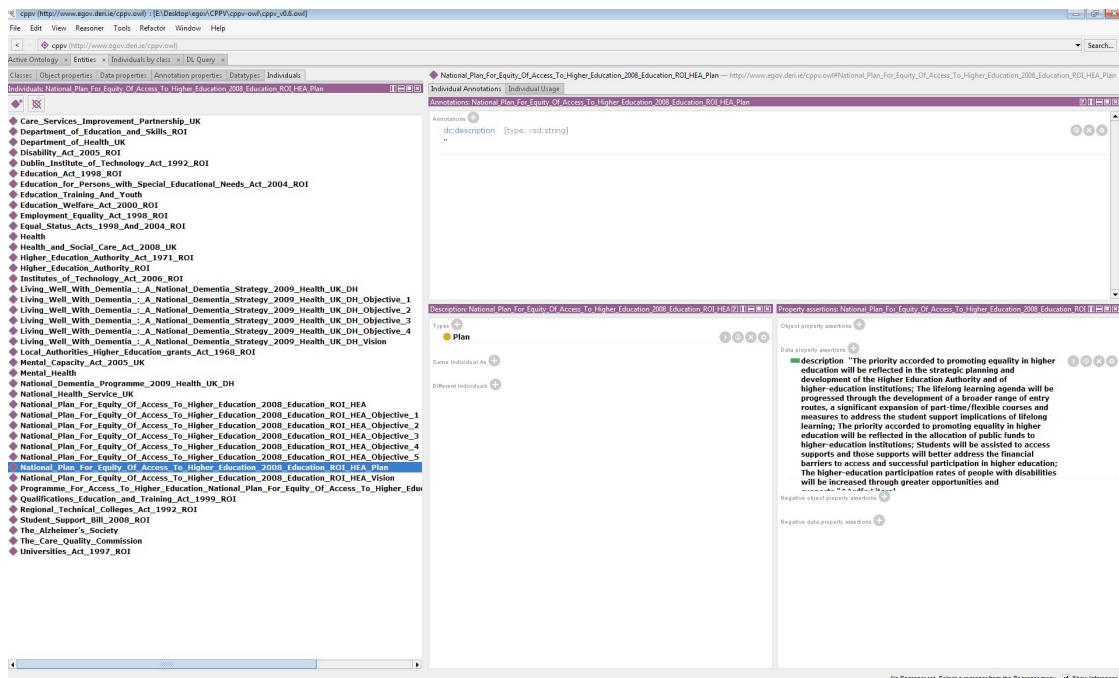


Figure 18 Protégé Implementation (CPPV Individuals metadata “sample 2”)

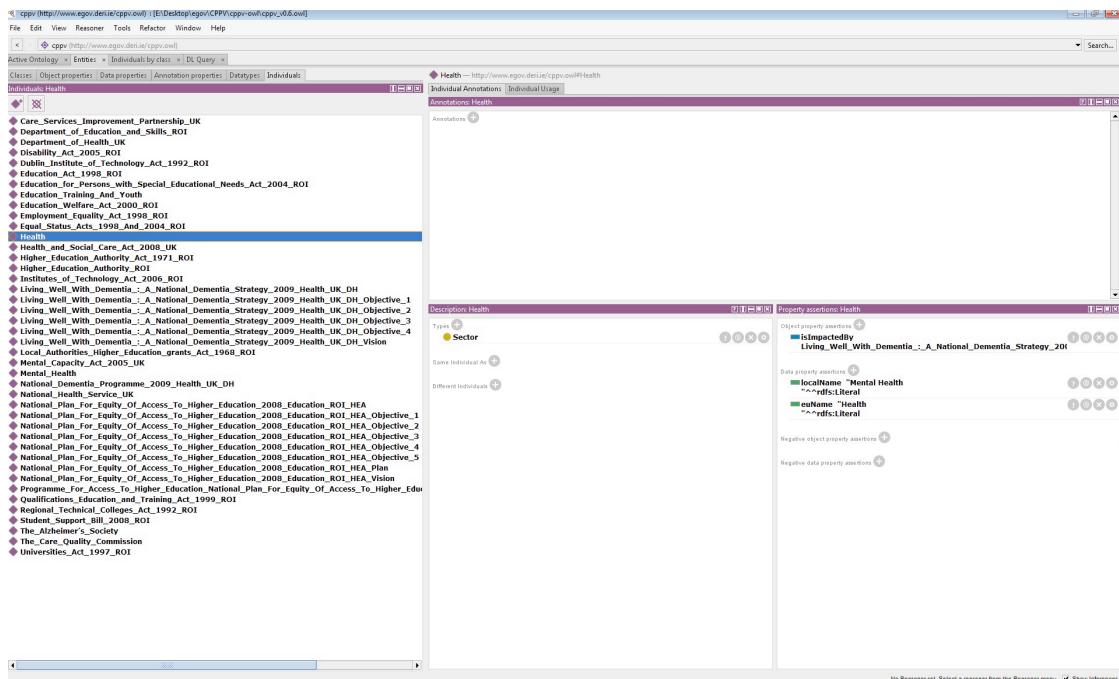


Figure 19 Protégé Implementation (CPPV Individuals metadata “sample 3”)

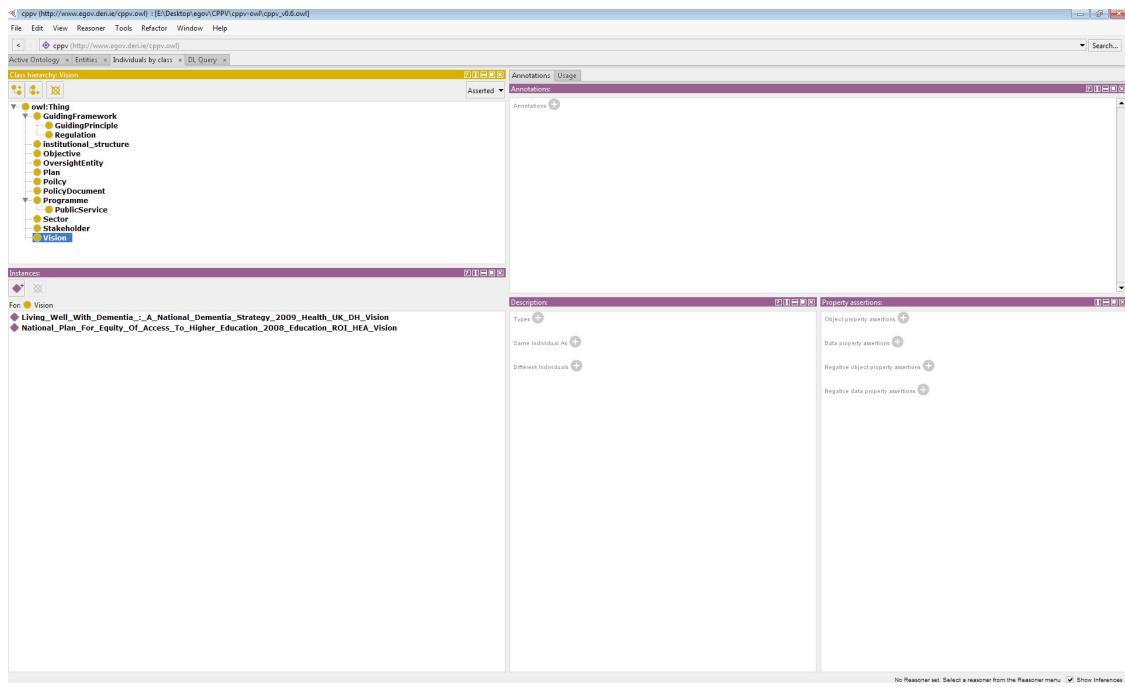


Figure 20 Protégé Implementation (CPPV Class Individuals “sample 1”)

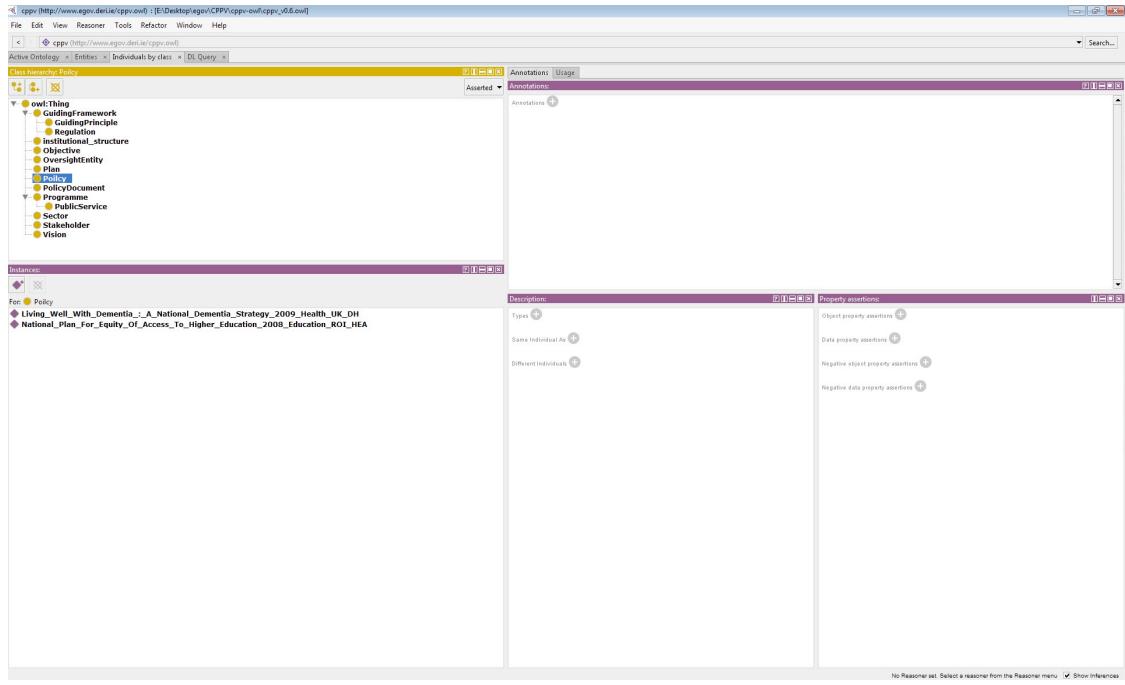


Figure 21 Protégé Implementation (CPPV Class Individuals “sample 2”)

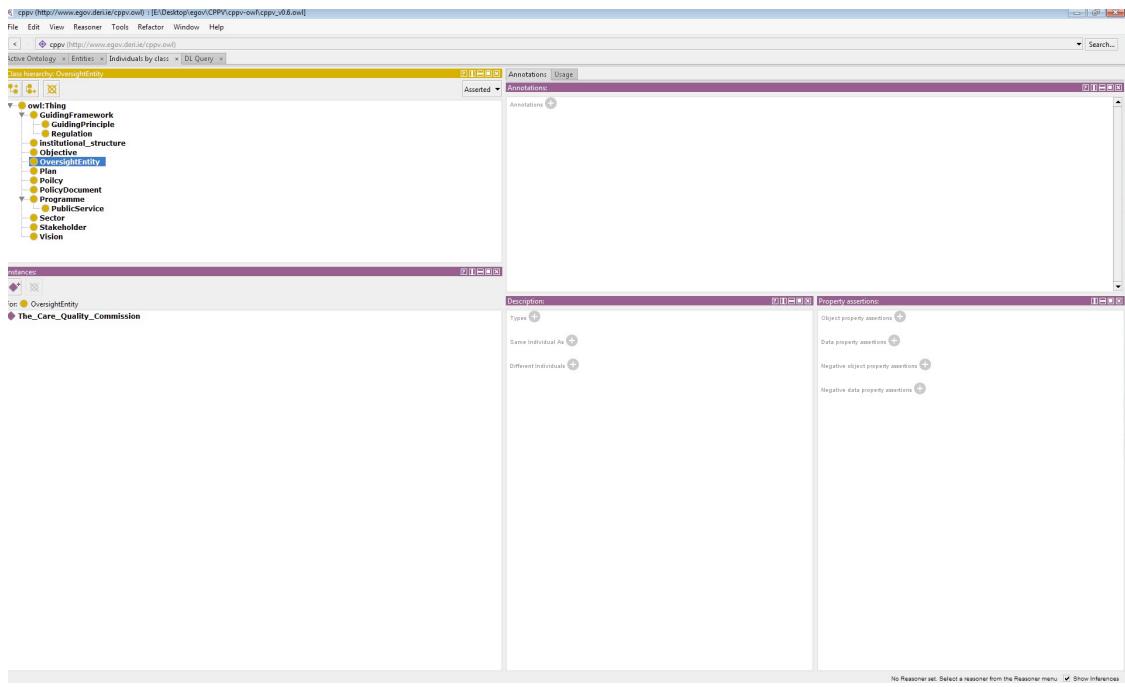


Figure 22 Protégé Implementation (CPPV Class Individuals “sample 3”)

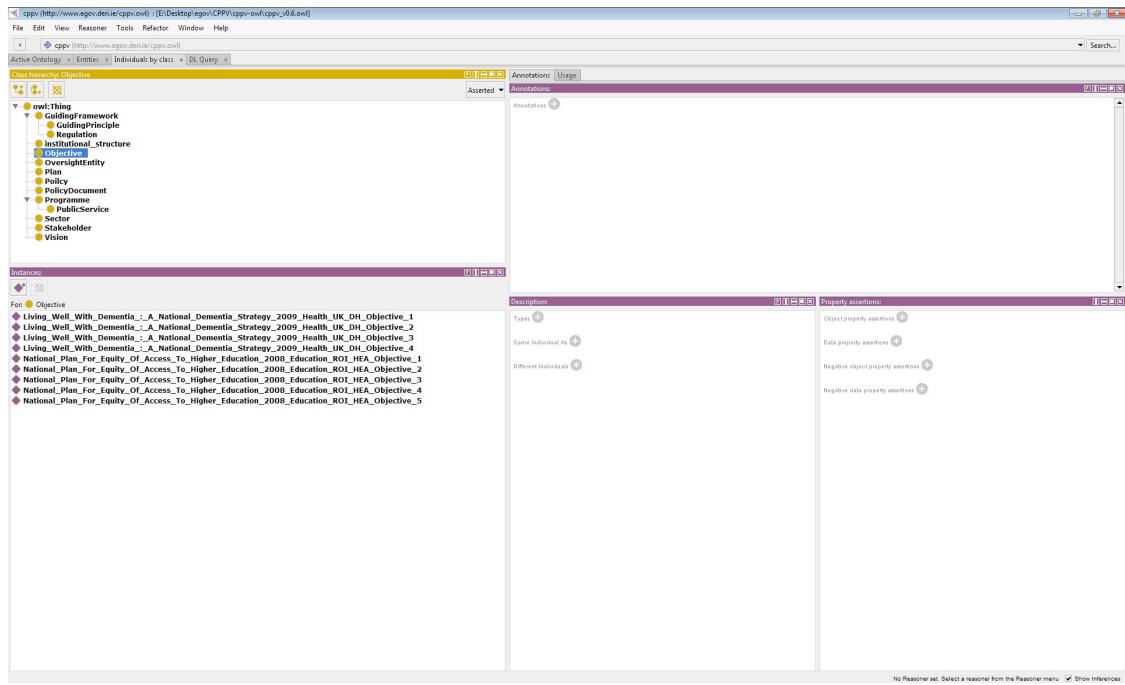


Figure 23 Protégé Implementation (CPPV Class Individuals “sample 4”)

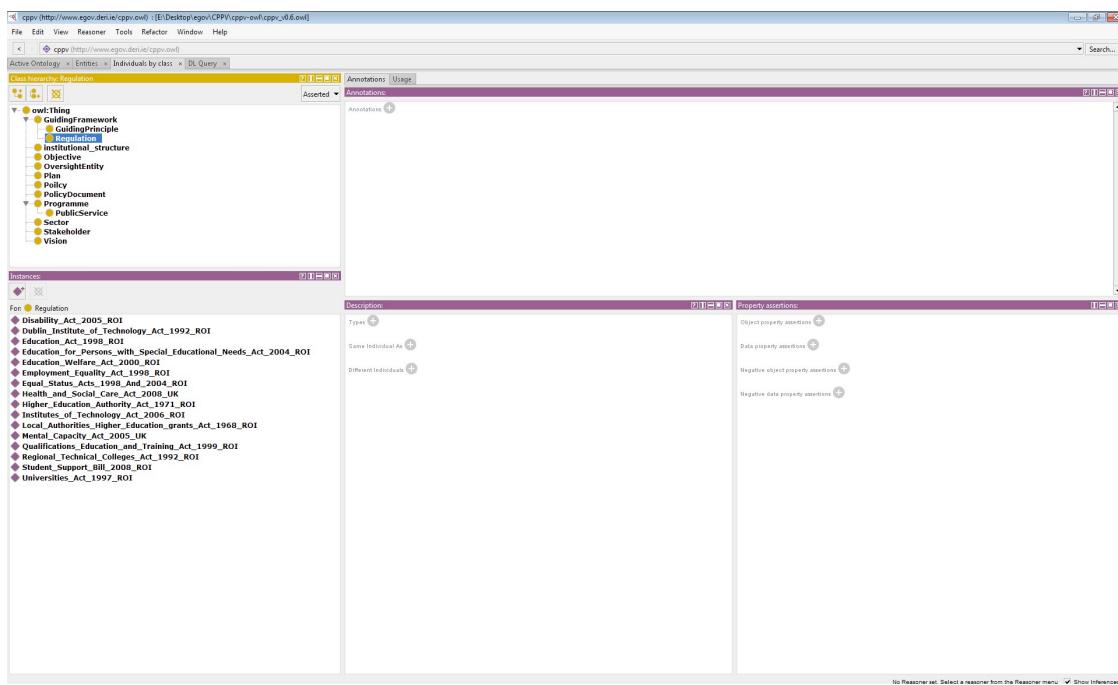


Figure 24 Protégé Implementation (CPPV Class Individuals “sample 5”)

4.1.9 CKAN Implementation

CPPV custom CKAN metadata json schema source code can be found at ¹⁰, live instance at ¹¹, and in appendix B. Following screen samples to illustrate the CPPV implementation in CKAN Figure 25 to Figure 28.

¹⁰ <https://github.com/mohadelrezk/cppv-ckan-implementation>

¹¹ <http://vmrtpa05.deri.ie:8080/dataset>

4.1.10 CKAN Screen Samples

Search Datasets... 🔍

17 datasets found

Order by: Relevance

Organizations
Insight Centre For ... (16)
The Department of H... (1)

Groups
There are no Groups that match this search

Tags
education (2)
Social welfare (2)
Health (2)
Communications (2)
youth (1)
trading online (1)
tourism (1)
sustainable protection (1)
social inclusion (1)
pensions (1)
[Show More Tags](#)

Formats
PDF (17)

Licenses
UK Open Government ... (1)
Creative Commons At... (1)

Environmental Policy Statement
Environmental Policy Statement
[PDF](#)

Ireland Broadband Intervention Strategy
Ireland Broadband Intervention Strategy
[PDF](#)

Irelands Transition to a Low Carbon Energy Future 2015-2030
Irelands Transition to a Low Carbon Energy Future 2015-2030
[PDF](#)

20-Year Strategy For The Irish Language 2010-2030
20-Year Strategy For The Irish Language 2010-2030
[PDF](#)

National Pensions Framework
National Pensions Framework
[PDF](#)

The Irish National Dementia Strategy
The Irish National Dementia Strategy
[PDF](#)

National Youth Strategy 2015 to 2020
National Youth Strategy 2015 to 2020
[PDF](#)

National Digital Strategy For Ireland
National Digital Strategy For Ireland
[PDF](#)

National Landscape Strategy
National Landscape Strategy
[PDF](#)

National Strategic Plan For Sustainable Aquaculture Development
National Strategic Plan For Sustainable Aquaculture Development
[PDF](#)

National Action Plan For Social Inclusion
National Action Plan For Social Inclusion
[PDF](#)

National Plan For Equity Of Access To Higher Education
National Plan For Equity Of Access To Higher Education
[PDF](#)

National Women's Strategy 2007 - 2016
National Women's Strategy 2007 - 2016
[PDF](#)

National Disability Strategy
National Disability Strategy
[PDF](#)

Smarter Travel - A Sustainable Transport Future
Smarter Travel - A Sustainable Transport Future
[PDF](#)

Tourism Policy
Tourism Policy
[PDF](#)

National Strategy for Action on Suicide Prevention 2005-2014
National Strategy for Action on Suicide Prevention 2005-2014
[PDF](#)

You can also access this registry using the [API](#) (see [API Docs](#)).

Figure 25 CPPV CKAN Implementation of Public Policies' Entries

Insight Search Datasets Organizations Groups About Log In Register EN

Organizations / Insight Centre For Data ... / 20-Year Strategy For The ...

Datasets Groups Activity Stream

20-Year Strategy For The Irish Language 2010-2030

20-Year Strategy For The Irish Language 2010-2030

Data and Resources

[20-Year-Strategy-for-the-Irish-Language-2010-2030.pdf](#)

department of education and skills - irish language

Additional Info

Field	Value
Policy Level	National
Policy Country	Ireland
Policy Language	English
Policy Start Date	2010-01-01
Policy Document Name	20-Year-Strategy-for-the-Irish-Language-2010-2030.pdf
Policy Document Description	20-Year-Strategy-for-the-Irish-Language-2010-2030.pdf
Policy Document File Type	pdf
Policy Document URL	http://www.education.ie/en/Publications/Policy-Reports/20-Year-Strategy-for-the-Irish-Language-2010-2030.pdf
Vision(s)	To create a supportive framework and the opportunities in which Irish can be passed on in a natural way within households.
Objective(s)	To increase the number of families throughout the country who speak Irish at home; To increase the number of people who provide linguistic support for the Gaeltacht as an Irish-speaking community; Ensure that Irish is the household and community language; Ensure that in the public sphere, Irish will be used as much as English will be, as far as practical, a choice for the citizen to make; Encourage the use of Irish in business; Encourage the State to choose to do their business in Irish; Ensure that Irish is more visible in our society, both as a spoken language by our citizens and also in areas such as signage and literature
Sector EU Name	Education
Sector Local Name	Education
Sector Description	Education
Oversight Entity Name(s)	Department of Education and Skills (DES)
OversightEntity Description	Department of Education and Skills (DES)
Programme(s)	20-year-strategy-for-the-irish-language-2010-2030
Plan	20-year-strategy-for-the-irish-language-2010-2030
Regulation Name(s)	Official Languages Act (2003), the Education Act (1996), Guidelines for the Irish Language Development Act (2000) and the Broadcasting Act (2001)
Guiding Regulation Description	Official Languages Act (2003), the Education Act (1996), Guidelines for the Irish Language Development Act (2000) and the Broadcasting Act (2001)
Guiding Principle Name(s)	
Guiding Principle Description	
Stakeholder Name(s)	Department of Arts, Heritage and the Gaeltacht (DAHG); Department of Education and Skills (DES); The National Council for Curriculum and Assessment (NCCA); Department of Children and Youth Affairs (DCYA); The Teaching Council; Department of Arts, Heritage and the Gaeltacht (DAHG); Department of Education and Skills (DES); The National Council for Curriculum and Assessment (NCCA); Department of Children and Youth Affairs (DCYA); The Teaching Council
Stakeholder Description(s)	
Stakeholder type(s)	
Institutional Structure Description	Department of Arts, Heritage, Gaeltacht (DAHG); Department of Education and Skills (DES)
Source	
Version	
Author	
Author Email	
Maintainer	
Maintainer Email	

Followers 0

Organization

Insight Centre For Data Analytics
Insight Centre For Data Analytics
eGovernment Unit, Core Public Policy
Vocabulary team read more

Social

- Facebook
- Twitter
- Google+

Creative Commons Attribution

Figure 26 CPPV CKAN Implementation of Public Policies' Entries (Public Policy sample 1)

Insight Search Datasets Organizations Groups Activity Stream

Organizations / Insight Centre For Data ... / Environmental Policy Statement

Datasets Groups Activity Stream

Environmental Policy Statement

Environmental Policy Statement

Data and Resources

[Environmental-Policy-Statement.pdf](#)

Custom House, Environment, Prevention of Pollution, energy consumption

Additional Info

Field	Value
Policy Level	National
Policy Country	Ireland
Policy Language	English
Policy Start Date	2015-01-01
Policy Document Name	FileDownload:16785.en
Policy Document Description	FileDownload:16785.en
Policy Document File Type	pdf
Policy Document URL	http://www.environ.ie/en/Environment/Publications/Documents/FileDownload:16785.en.pdf
Vision(s)	The achievement of sustainable development
Objective(s)	Prevention of pollution; Continue to decrease energy consumption in the Custom House
Sector EU Name	Environment
Sector Local Name	Environment
Sector Description	Environment
Oversight Entity Name(s)	Gross-Divisional Environment Management Team
OversightEntity Description	Gross-Divisional Environment Management Team
Programme(s)	The Environmental Management Programme 2015/2016
Plan	Protect the environment; Manage resources and promote sustainable growth; Support citizens for safe and sustainable communities
Regulation Name(s)	Complies with all relevant environmental legislation and standards
Guiding Regulation Description	Complies with all relevant environmental legislation and standards
Guiding Principle Name(s)	
Guiding Principle Description	
Stakeholder Name(s)	Department of the Environment, Community and Local Government
Stakeholder Description(s)	Department of the Environment, Community and Local Government
Stakeholder type(s)	
Institutional Structure Description	
Source	
Version	
Author	
Author Email	
Maintainer	
Maintainer Email	

Followers 0

Organization

Insight Centre For Data Analytics
Insight Centre For Data Analytics
eGovernment Unit, Core Public Policy
Vocabulary team read more

Social

- Facebook
- Twitter
- Google+

Figure 27 CPPV CKAN Implementation of Public Policies' Entries (Public Policy sample 2)

Search Datasets Organizations Groups About Log In Register EN

Organizations / Insight Centre For Data ... / Irelands Transition to a Low ...

Dataset Groups Activity Stream

Irelands Transition to a Low Carbon Energy Future 2015-2030

Irelands Transition to a Low Carbon Energy Future 2015-2030

Data and Resources

[Energy White Paper - Dec 2015.pdf](#) Energy White Paper - Dec 2015.pdf

[Explore](#)

Communications SEAI programmes emissions energy

Additional Info

Field	Value
Policy Level	National
Policy Country	Ireland
Policy Language	English
Policy Start Date	2015-01-01
Policy Document Name	Energy White Paper - Dec 2015.pdf
Policy Document Description	Energy White Paper - Dec 2015.pdf
Policy Document File Type	pdf
Policy Document URL	http://www.dccae.gov.ie/energy/Lists/Publications/2015Documents/Energy%20White%20Paper%20-%20Dec%202015.pdf
Vision(s)	To transform Ireland into a low carbon society and economy by 2050
Objective(s)	Reduce GHG emissions from the elie ensuring sustainability, security of supply and competitiveness in the energy sector
Sector EU Name	Energy
Sector Local Name	Energy
Sector Description	Energy
Oversight Entity Name(s)	The SEAI Public Sector Programme
OversightEntity Description	The SEAI Public Sector Programme.
Programme(s)	EU emissions trading scheme (ETS); France-United Kingdom-Ireland (FUI) Regional Initiative; The National Smart Metering Programme; The Government's Electric Vehicle (EV) Grant Scheme; SEAI Better Energy Programmes; SEAI Schools Programme
Plan	Replace fossil fuels with renewable energy sources.
Regulation Name(s)	The Climate Action and Low Carbon Development Bill 2015; EU Directive 2009/72/EC [55]
Guiding Regulation Description	The Climate Action and Low Carbon Development Bill 2015; EU Directive 2009/72/EC [55]
Guiding Principle Name(s)	The citizen will be at the centre of the transition and the energy industry. Government and public authorities will all contribute to ensuring that citizens are involved in the transition, and benefit from it. The transition will be underpinned by policy and regulation that is stable and predictable. Policy will seek to achieve maximum benefit at least cost, while reducing decarbonisation will both incur costs and deliver benefits; Policy measures will be evidence based and subject to rigorous analysis and appraisal prior to being implemented. This appraisal will include rigorous analysis of the distributional impacts of policy measures; New technologies, services, energy network architectures and community-based energy initiatives will emerge to facilitate and drive the transition. Policy will evolve to optimise the benefits from innovative initiatives.
Guiding Principle Description	The citizen will be at the centre of the transition and the energy industry. Government and public authorities will all contribute to ensuring that citizens are involved in the transition, and benefit from it. The transition will be underpinned by policy and regulation that is stable and predictable. Policy will seek to achieve maximum benefit at least cost, while reducing decarbonisation will both incur costs and deliver benefits; Policy measures will be evidence based and subject to rigorous analysis and appraisal prior to being implemented. This appraisal will include rigorous analysis of the distributional impacts of policy measures; New technologies, services, energy network architectures and community-based energy initiatives will emerge to facilitate and drive the transition. Policy will evolve to optimise the benefits from innovative initiatives.
Stakeholder Name(s)	Department of Communications, Energy and Natural Resources (DCCENR); The Energy Research Strategy Group (ERSG); The Department of the Environment, Community and Local Government (DECLG); Economic and Social Research Institute (ESRI); The Environmental Research Unit of University College Cork (UCC); Environmental Protection Agency (EPA); Department of Trade and Investment; Department of Education and Skills (DES)
Stakeholder Description(s)	Department of Communications, Energy and Natural Resources (DCCENR); The Energy Research Strategy Group (ERSG); The Department of the Environment, Community and Local Government (DECLG); Economic and Social Research Institute (ESRI); The Environmental Research Unit of University College Cork (UCC); Environmental Protection Agency (EPA); Department of Trade and Investment; Department of Education and Skills (DES)
Stakeholder type(s)	
Institutional Structure Description	Department Of Communications, Energy And Natural Resources; SEAI programmes
Source	
Version	
Author	
Author Email	
Maintainer	
Maintainer Email	

Irelands Transition to a Low C carbon Energy Future 2015-2030

Followers 0

Organization

Insight Centre For Data Analytics
Insight Centre For Data Analytics, e-Government Unit, Core Public Policy Vocabulary team [read more](#)

Social

Google+ Twitter Facebook

Figure 28 CPPV CKAN Implementation of Public Policies' Entries (Public Policy sample 3)

4.2 Public Policy Collaboration Mining for Citizen Satisfaction

4.2.1 Semantic Profiling for Collaboration Mining

As shown in Figure 4 and zoomed in Figure 29 we have designed a solution pipeline that incorporates Data Acquisition, Data Modelling, Data Analysis, and Data visualization technologies to enable the existence of a collaboration mining tool. We start with inputting the targeted open data portal(s) in which we seek mining for collaborations then we start acquiring metadata (catalogue) of the datasets. Then we restructure the catalogue to fit into the predesigned storage model (semantic profile), within this model we enhance, filter and exclude less important catalogue features – regarding our use case - and we add new features that are corresponding to our collaboration mining requirements e.g. we add “textual representation” feature by merging original textual features of the data catalogue, we add “Entities” feature to the new catalogue storage model by applying NER over the new “textual representation” feature of the catalogue, we filter features like “author” and “creator” to end up with only “publisher ID” feature, and we exclude “groups” and “tracking summary” features as showing in Figure 30 and Figure 31. After constructing and storing the new data model (semantic profile) we start the unstructured data analysis (text mining) pipeline by applying NER algorithm. At the end of that process we generate a comparable feature “Entities” and add it to the new data model to be used for collaboration mining. After that we construct dataset’s publisher data model (semantic profiles) which contains aggregated features’ values from their published datasets. Finally, we compute relation strengths between dataset publishers based on *comparing* their semantic profiles that we built using the aggregation of unique entities they publish datasets about and store it as shown in Figure 32 for later visualization and web service usages as shown in Figure 35 and Figure 36.

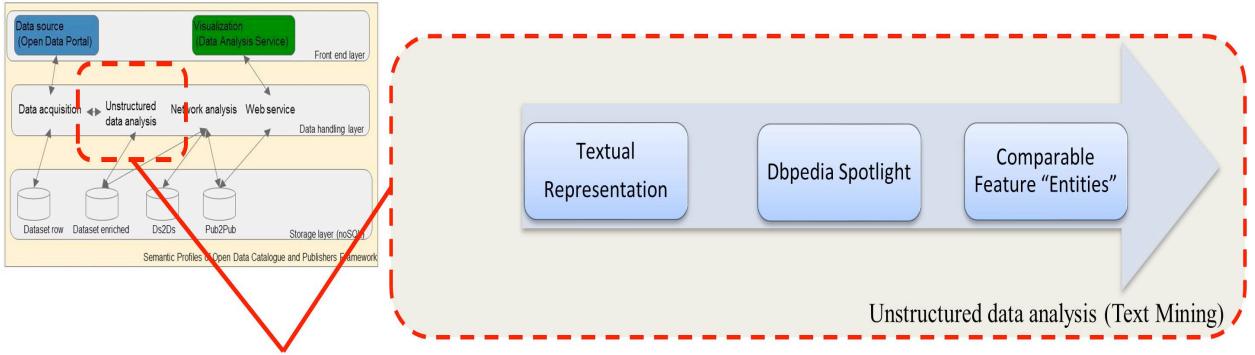


Figure 29 Unstructured Data Analysis (Text Mining)

Key	Value	Type	Key	Value	Type
(1) ocean-energy-related.portal	{ \$ (1) }	Object	[1] zoning_plan	zoning_plan	String
↳ dataset	↳ dataset	Object	↳ local_development	local_development	String
↳ dataset_id	↳ dataset_id	Object	↳ plan_statutory	plan_statutory	String
↳ dataset_modified	↳ dataset_modified	Object	↳ Department_Housing_Planning	Department_Housing_Planning	String
↳ dataset_name	↳ dataset_name	Object	↳ Array [2]	Array [2]	Object
↳ dataset_type	↳ dataset_type	Object	↳ type	type	String
↳ dataset_version	↳ dataset_version	Object	↳ entity	entity	Object
↳ dataset_over_id	↳ dataset_over_id	Object	[1] ↳ type	type	String
↳ dataset_over_modified	↳ dataset_over_modified	Object	[1] ↳ entity	entity	Object
↳ dataset_over_name	↳ dataset_over_name	Object	[2] ↳ type	type	String
↳ dataset_over_type	↳ dataset_over_type	Object	[2] ↳ entity	entity	Object
↳ dataset_over_version	↳ dataset_over_version	Object	↳ Stanford_Entities	Stanford_Entities	Object
↳ dataset_over_modified	↳ dataset_over_modified	Object	↳ rdfData	rdfData	Object
↳ dataset_over_name	↳ dataset_over_name	Object	↳ DBpedia_Refined	DBpedia_Refined	Object
↳ dataset_over_type	↳ dataset_over_type	Object	↳ DBpedia_Thing	DBpedia_Thing	Object
↳ dataset_over_version	↳ dataset_over_version	Object	↳ wave_energy	wave_energy	String
↳ dataset_over_modified	↳ dataset_over_modified	Object	[1] ↳ zoning_plan	zoning_plan	String
↳ dataset_over_name	↳ dataset_over_name	Object	[1] ↳ local_development	local_development	String
↳ dataset_over_type	↳ dataset_over_type	Object	[1] ↳ plan_statutory	plan_statutory	String
↳ dataset_over_version	↳ dataset_over_version	Object	[1] ↳ Department_Housing_Planning	Department_Housing_Planning	String
↳ dataset_over_modified	↳ dataset_over_modified	Object	[1] ↳ zoning_plan	zoning_plan	String
↳ dataset_over_name	↳ dataset_over_name	Object	[2] ↳ 15-years-and-over-and-present-in-the-state-number-by-district-type	15-years-and-over-and-present-in-the-state-number-by-district-type	String
↳ dataset_over_type	↳ dataset_over_type	Object	[2] ↳ 15-years-and-over-and-present-in-the-state-number-by-regional-authority-detailed-type	15-years-and-over-and-present-in-the-state-number-by-regional-authority-detailed-type	String
↳ dataset_over_version	↳ dataset_over_version	Object	[3] ↳ 15-years-and-over-and-present-in-the-state-number-by-regional-authority-detailed-type	15-years-and-over-and-present-in-the-state-number-by-regional-authority-detailed-type	String

Figure 30 Original Dataset Metadata (Catalogue) and Collaboration Mining Ready Data Models

Key	Value	Type	Key	Value	Type
(1) Gillian_Gannon-smartbay-ireland	{ \$ (1) }	Object	[1] zoning_plan	zoning_plan	String
↳ id	Gillian_Gannon-smartbay-ireland	String	↳ local_development	local_development	String
↳ commonEntities	Array [3]	Array	↳ plan_statutory	plan_statutory	String
↳ type	DisplayMatching	Object	↳ Department_Housing_Planning	Department_Housing_Planning	String
↳ entity	renewable_energy	String	↳ Array [2]	Array [2]	Object
[1] ↳ type	DisplayMatching	Object	[1] ↳ type	type	String
[1] ↳ entity	ocean_energy	String	[1] ↳ entity	entity	Object
[2] ↳ type	DisplayMatching	Object	[2] ↳ type	type	String
[2] ↳ entity	ocean_energy	String	[2] ↳ entity	entity	Object
[3] ↳ type	DBpedia_Thing	Object	↳ Stanford_Entities	Stanford_Entities	Object
[3] ↳ entity	wave_energy	String	↳ rdfData	rdfData	Object
↳ relationStrength	3	Int32	↳ DBpedia_Refined	DBpedia_Refined	Object
↳ publisherA	smartbay-ireland	String	↳ DBpedia_Thing	DBpedia_Thing	Object
↳ publisherB		String	[1] ↳ zoning_plan	zoning_plan	String
↳ datasetA		String	[2] ↳ 15-years-and-over-and-present-in-the-state-number-by-district-type	15-years-and-over-and-present-in-the-state-number-by-district-type	String
↳ datasetB		String	[3] ↳ 15-years-and-over-and-present-in-the-state-number-by-regional-authority-detailed-type	15-years-and-over-and-present-in-the-state-number-by-regional-authority-detailed-type	String
↳ vectorA		String			
[0]	Cork	String			
[1]	celtic sea	String			
[2]	cork harbour	String			
[3]	Ireland	String			
[4]	ADCOP	String			
[5]	CSV	String			
[6]	acoustic doppler current profiler	String			
[7]	buoys	String			
[8]	hydrologic cycle	String			
[9]	hydrosphere	String			
[10]	ocean circulation	String			
[11]	ocean current	String			
[12]	oceanographic	String			
[13]	oceanography	String			
[14]	water science	String			
[15]	SM	String			
[16]	chemistry	String			
[17]	environmental monitoring	String			
[18]	environmental science	String			
[19]	recorder	String			
[20]	salinity	String			
[21]	science	String			
[22]	water quality	String			
[23]	JSON	String			
[24]	KML	String			

Figure 31 Publisher Collaboration Network

Following we discuss and represent our Semantic Profiling for Collaboration Mining approach results.

4.2.2 Profiling the Catalogues

By querying the stored enriched metadata of open data portal we are able to generate charts that are profiling the underlying open data catalogue. As an example of those queries we are able to retrieve the named entities detected from mining unstructured textual representations of data catalogues generated by our tool. Those

named entities which are originally derive from dataset metadata are - same as their origin – able to demonstrate a description of the contents of the data portals see Figure 32 and Figure 33.

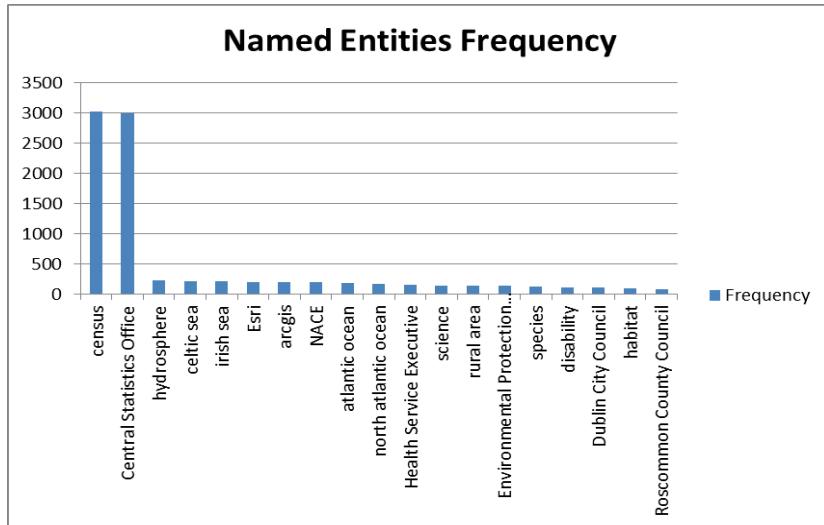


Figure 32 Top Named Entities Describing the Open Data Portal “data.gov.ie”

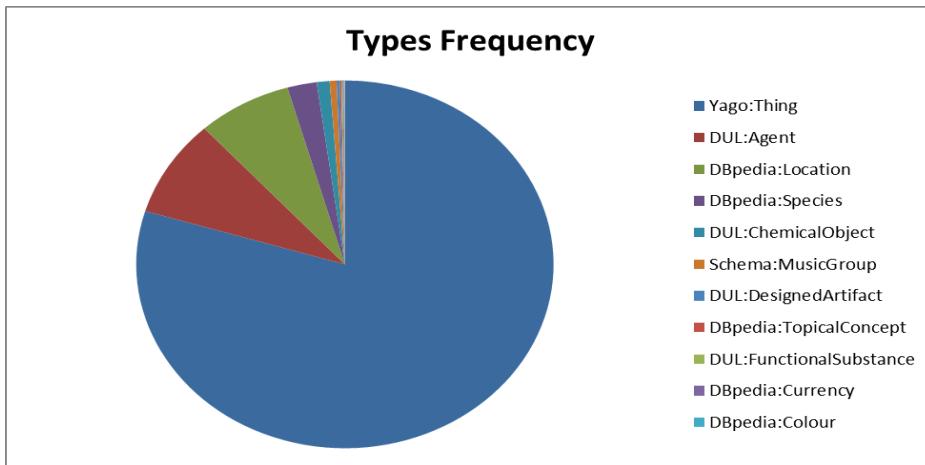


Figure 33 Top Named Entities Types Describing the Open Data Portal “data.gov.ie”

Publishers Profiles

Open data publishers are an interesting open data analysis feature; publishers could be governmental departments, councils, etc. which make their profiles a key component of data integration and standardization. An open data publisher's profile is the aggregation of the information extracted from its published dataset metadata. One of the usages of a publisher's profile is to understand more about the domain of the publisher see Figure 34 for an example.

Open Data Publisher	Named Entities;Frequency
central-statistics-office	Central Statistics Office;2985, census;2863, county;269, NACE;194, Irish;142, rural area;140, disability;97, Irish Travellers;68
department-of-housing-planning-community-and-local-government	department;209, Helena;96, Constitution;38, referendum;35, local authority;35, census;21, county;18, mortgage;15, gener
health-service-executive	Health Service Executive;140, achi;73, id;10;73, Discharge;62, DRG;62, AM;62, AR;60, Irish;59, ACS;56, birthweight;54, gest
environmental-protection-agency	Environmental Protection Agency;139, REST;102, WMS;74, Water Framework Directive;73, Informatics;59, WFD;43, Irish se
marine-institute	hydrosphere;119, atlantic ocean;91, north atlantic ocean;87, celtic sea;85, irish sea;85, CSV;61, biology;56, life science;54,
dublin-city-council	Dublin City Council;105, Dublin City;60, .csv;49, Irish;20, ID;18, Dublin;16, DCC;14, MAP;10, DAT;10, urban planning;9, recycl
roscommon-county-council	KML;85, CSV;85, Roscommon County Council;85, Shapefile;81, Esri;80, Rest API;80, arcgis;80, Roscommon;65, roscommon;2
geological-survey-of-ireland	geological survey;58, earth science;55, GSI;53, science;49, lithosphere;43, ireland;38, Irish;32, WMS;29, JSON;28, bedrock;2
c2f170ca-63d0-4498-9e81-759827708e97	Ordnance Survey;57, Shapefile;56, CSV;56, KML;56, Rest API;56, Esri;56, arcgis;55, OSi;55, level crossing;3, watercourse;2, S
national-transport-authority	statute;56, Dublin Transport Authority;56, Oireachtas;56, National Transport Authority;56, public transport;55, Peter;54, Di
all-island-research-observatory	demography;55, CSV;43, REST;30, JSON;30, SA;22, CSO;19, Service economy;14, population density;13, ED;12, Northern Cro
fingal-county-council	Fingal County Council;53, county;52, .csv;35, Fingal;24, WGS84;20, .zip;5, ID;4, datasheet;3, water quality;2, recycling;2, Du
department-of-education-and-skills	partment;50, census;40, county;11, Leaving Certificate;3, .csv;3, ireland;2, national school;2, Junior Certificate;2, xls;2, I
revenue-commissioners	Revenue Commissioners;47, income tax;8, Wicklow;7, Kerry;7, Cavan;7, Monaghan;7, Longford;7, Meath;7, Scheme;7, Kilk
health-research-board	census;42, ICD 10;12, Health Service Executive;9, group sex;3, database;2, disability;2, fibrosis;1, scooters;1, lumbar;1, inn
galway-county-council	Shapefile;42, CSV;42, Galway County Council;42, KML;41, Rest API;36, Esri;36, arcgis;36, county;35, Galway County;19, Galw
national-biodiversity-data-centre	wildlife protection;40, wildlife sanctuary;40, species;40, data centre;40, syncology;40, wildlife conservation;40, wildlife;
sustainable-energy-authority-of-ireland	ireland;38, renewable energies;26, renewable energy;26, Atlas;23, WMS;19, JSON;18, celtic sea;18, irish sea;18, north atla
dun-laoghaire-rathdown-county-council	DLR;18, .csv;7, sbn;6, shp;6, dbf;6, Cycling;5, st;3, ards;3, xls;3, Cherrywood;3, Sandyford;2, WW1;2, A2;2, Dev;2, A1;2, INFO
department-of-communications-climate-action-and-environment	department;36, hydrosphere;35, KML;32, CSV;31, WMS;30, WFS;29, north atlantic ocean;26, atlantic ocean;26, celtic sea;26

Figure 34 Top Named Entities Describing Data Posted by Top Publishers to the Open Data Portal “data.gov.ie”

4.2.3 Interlinking Publishers

The resulted publisher profiles are used to mine possible collaboration channels between data publishers at data portal level and among portals level using the added comparable feature (Entities) see Figure 35 to Figure 37.

According to our trial run “marine-institute (129) datasets” and “geological-survey-of-ireland (67) datasets” have the highest relation strength score of (82) which means that they share 82 entities in common. We examined the datasets published by both publishers and we found that for pollution concept there are (7) datasets published by “marine-institute” and (7) dataset published by “geological-survey-of-ireland” and similarly for hydrography concept there are (4) datasets published by “marine-institute” and (18) datasets published by “geological-survey-of-ireland” as shown in Figure 35 and Figure 36.

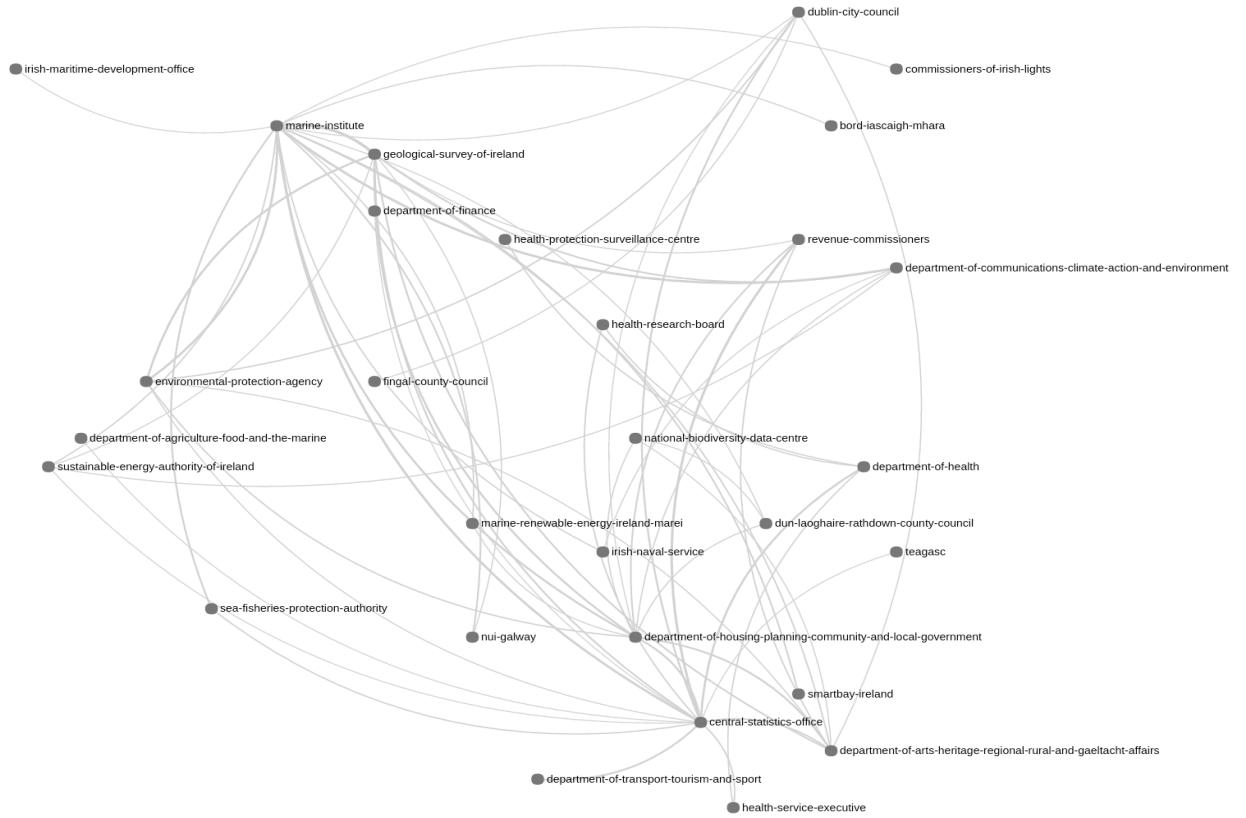


Figure 35 Publishers Collaboration Network of Open Data Portal “data.gov.ie” with relation strength > 20

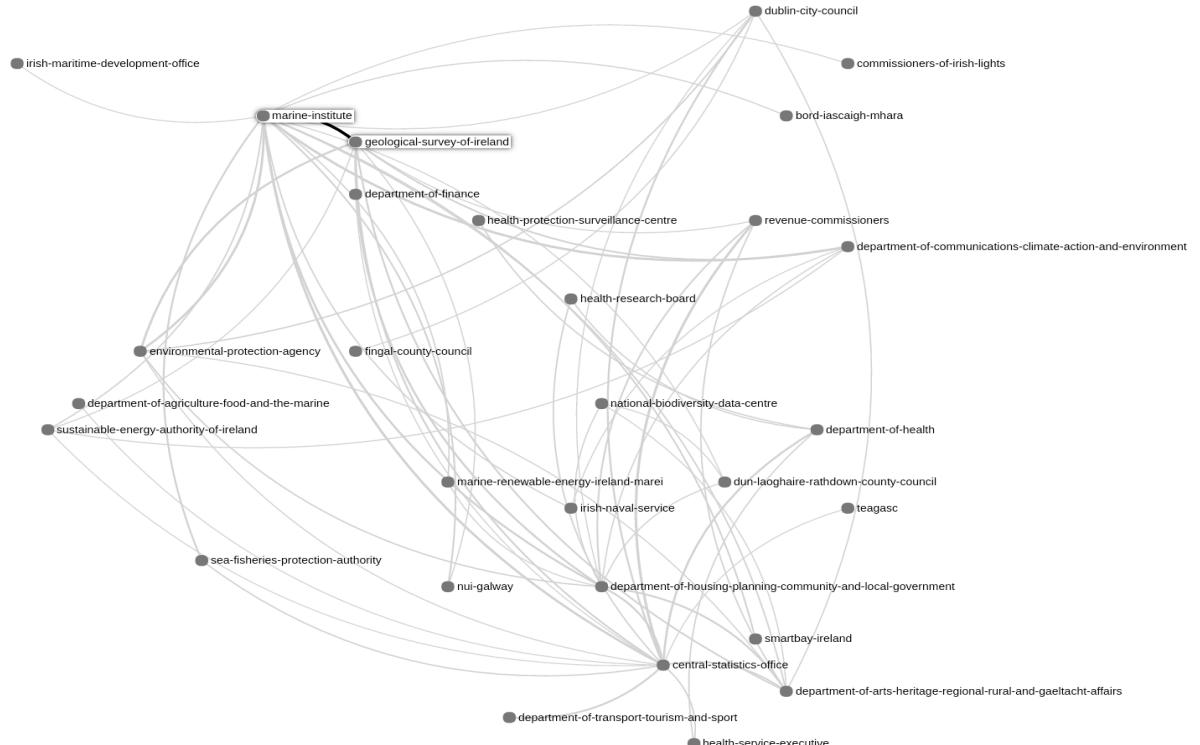


Figure 36 Publishers Collaboration Network of Open Data Portal “data.gov.ie” – Showing highest relation strength score between “marine-institute” and “geological-survey-of-ireland”

publisherA	publisherB	strength	commons
marine-institute	geological-survey-of-ireland	82	[("Dbpedia:Thing", "pollution"), ("Dbpedia:Thing", "hydrography"), ("Dbpedia:Thing", "habitat"), ("Dbpedia:Location", "United Nations"), ("Dbpedia:Thing", "hydrography"), ("Dbpedia:Thing", "North Atlantic Ocean"), ("Dbpedia:Thing", "habitat"), ("Dbpedia:Species", "whiting"), ("Dbpedia:Thing", "probability"), ("Dbpedia:Thing", "Stamp Duty"), ("Dbpedia:Thing", "Exchequer"), ("Dbpedia:Thing", "communication"), ("Dbpedia:Thing", "aquatic"), ("Dbpedia:Thing", "hydrography"), ("Dbpedia:Thing", "estuary"), ("Dbpedia:Thing", "topography"), ("Dbpedia:Thing", "hydrography"), ("Dbpedia:Thing", "habitat"), ("Dbpedia:Thing", "density"), ("Dbpedia:Thing", "Shapefile"), ("Dbpedia:Thing", "birthweight"), ("Wikidata:Q12136", "brucellosis"), ("Wikidata:Q12136", "influenza"), ("Dbpedia:Thing", "birthweight"), ("Dbpedia:Thing", "habitat"), ("Dbpedia:Location", "Fingal"), ("Dbpedia:Thing", "Shapefile"), ("Dbpedia:Thing", "county"), ("Dbpedia:Thing", "wildlife protection"), ("Dbpedia:Thing", "habitat"), ("Dbpedia:Thing", "wildlife conservation"), ("Dbpedia:Thing", "tonne"), ("Dbpedia:Thing", "road safety"), ("Dbpedia:Thing", "force"), ("Dbpedia:Thing", "copper"), ("Dbpedia:Thing", "aluminum"), ("Dbpedia:Thing", "force"), ("Dbpedia:Thing", "geophysical"), ("Dbpedia:Thing", "hydrography"), ("Dbpedia:Thing", "Shapefile"), ("Dbpedia:Thing", "habitat"), ("Dbpedia:Species", "whiting"), ("Dbpedia:Thing", "fish"), ("Dbpedia:Location", "Sligo"), ("Dbpedia:Thing", "habitat"), ("Dbpedia:Location", "Vauxhall"), ("DUL:Agent", "piaggio"), ("DUL:Agent", "skoda")]
department-of-communications-climate-action-and-environment	marine-institute	66	
central-statistics-office	marine-institute	62	
central-statistics-office	revenue-commissioners	62	
marine-institute	environmental-protection-agency	58	
environmental-protection-agency	geological-survey-of-ireland	55	
department-of-housing-planning-community-and-local-government	marine-institute	53	
central-statistics-office	department-of-health	51	
department-of-housing-planning-community-and-local-government	geological-survey-of-ireland	50	
department-of-housing-planning-community-and-local-government	central-statistics-office	48	
department-of-housing-planning-community-and-local-government	department-of-arts-heritage-regional-rural-and-gaeltacht-affairs	48	
central-statistics-office	department-of-transport-tourism-and-sport	47	
central-statistics-office	dublin-city-council	43	
central-statistics-office	geological-survey-of-ireland	42	
department-of-communications-climate-action-and-environment	geological-survey-of-ireland	42	
sea-fisheries-protection-authority	marine-institute	42	
department-of-arts-heritage-regional-rural-and-gaeltacht-affairs	geological-survey-of-ireland	41	

Figure 37 Publishers Mined Relations of Open Data Portal “data.gov.ie”

4.3 GOV DAF

4.3.1 Topic Modelling for Public Policy Analysis Module

To test Gov-DAF knowledge base building pipeline accuracy, we applied the aforementioned methodologies over two different public policy sample documents (UK Bioenergy Strategy and Irish National Plan for Equity of Access to Higher Education [162], [163]) for keywords extraction, and manually extracting public policy aspects i.e. objectives. Then relating keywords back to objectives and using human evaluation of the results we calculated the following performance measures used in machine learning domain [164] Equation 2.

As shown in Table 8 and Figure 3840, Topic Modelling Methodology appears to prove higher accuracy indices compared with Named Entity Recognition Methodology as it have higher Accuracy, Precision, Recall, F measure and lower Error rate. Depending on those results Gov-DAF knowledge base pipeline will adopt and enhance Topic Modelling Methodology for the last phases of the research.

Equation 2 Accuracy Measures

$$\text{Classification Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Error Rate} = \frac{FP + FN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$f \text{ measure} = \frac{2PR}{P + R}$$

Table 8 Accuracy Measures Results Tabular View

	NER		Topic Modeling	
	PolicyA	PolicyB	PolicyA	PolicyB
Accuracy	0.3	0.51	0.82	0.74
Error Rate	0.7	0.49	0.18	0.26
F measure	0	0	0.75	0.62
Precision	0	0	1	1
Recall	0	0	0.6	0.45

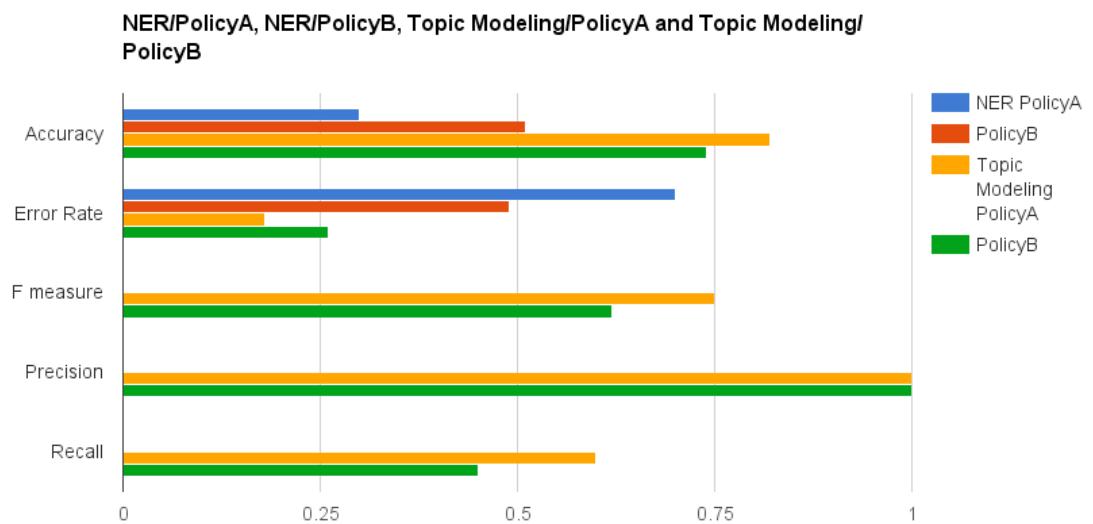
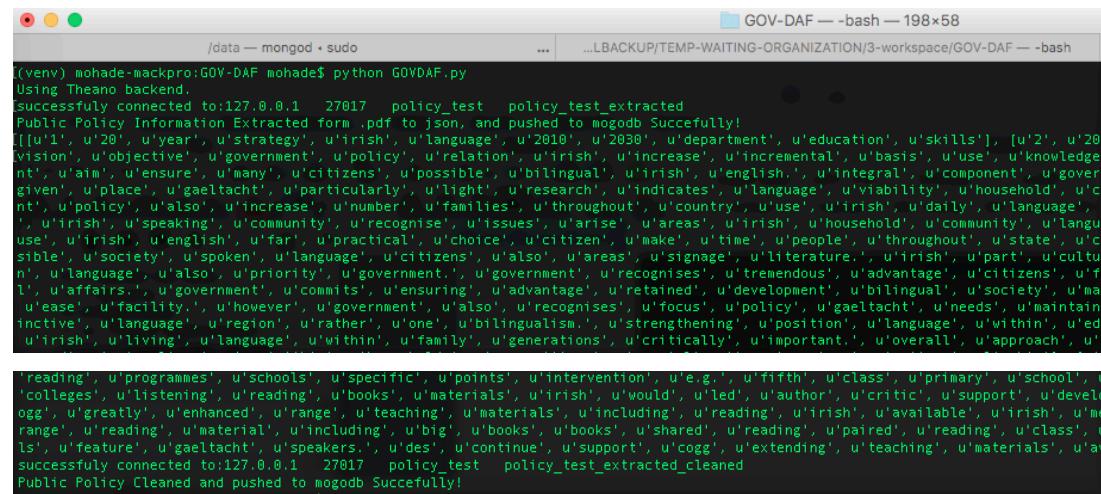


Figure 38 Accuracy Measures Compared

4.3.2 Citizen Satisfaction Index Estimation Pipeline

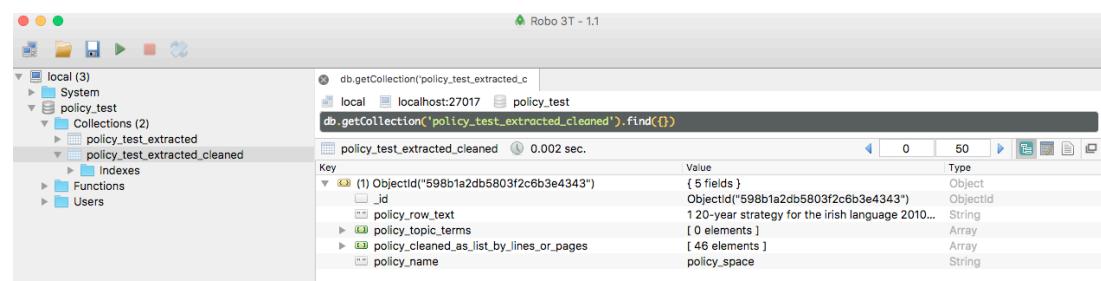
Citizen Satisfaction Index Analysis Pipeline shown in Figure 1. Source code can be found at ¹², and in appendix C. Following the experiment screen samples of the Citizen Satisfaction Analysis Pipeline's different modules in Figure 3941 to Figure 5456 and in Equation 3:

4.3.3 Public Policy Data Extraction and Cleansing Module



```
(venv) mohade-mackpro:GOV-DAF mohade$ python GOVDAF.py
Using Theano backend.
Successfully connected to:127.0.0.1 27017 policy_test policy_test_extracted
Public Policy Information Extracted from .pdf to json, and pushed to mongodb Successfully!
[{"u'1', u'20', u'year', u'strategy', u'irish', u'language', u'2010', u'2030', u'department', u'education', u'skills'], [u'2', u'vision', u'objective', u'government', u'policy', u'relation', u'irish', u'increase', u'incremental', u'basis', u'use', u'knowledge', u'nt', u'aim', u'ensure', u'many', u'citizens', u'possible', u'bilingual', u'irish', u'english', u'integral', u'component', u'given', u'place', u'gaeltacht', u'particularly', u'light', u'research', u'indicates', u'language', u'visibility', u'household', u'cont', u'policy', u'also', u'increase', u'number', u'families', u'throughout', u'country', u'use', u'irish', u'daily', u'language', u', u'irish', u'speaking', u'community', u'recognise', u'issues', u'arise', u'areas', u'irish', u'household', u'community', u'language', u'use', u'irish', u'english', u'far', u'practical', u'choice', u'citizen', u'make', u'time', u'people', u'throughout', u'state', u'ch', u'society', u'spoken', u'language', u'citizens', u'also', u'areas', u'signage', u'literature', u'irish', u'part', u'culture', u'language', u'also', u'priority', u'government', u'government', u'recognises', u'tremendous', u'advantage', u'citizens', u'f', u'affairs', u'government', u'commits', u'ensuring', u'advantage', u'retained', u'development', u'bilingual', u'society', u'man', u'ease', u'facility', u'however', u'government', u'also', u'recognises', u'focus', u'policy', u'gaeltacht', u'needs', u'maintain', u'active', u'language', u'region', u'rather', u'one', u'bilingualism', u'strengthening', u'position', u'language', u'within', u'edu', u'irish', u'living', u'region', u'within', u'family', u'generations', u'critically', u'important', u'overall', u'approach', u'reading', u'programmes', u'schools', u'specific', u'points', u'intervention', u'e.g.', u'fifth', u'class', u'primary', u'school', u'colleges', u'listening', u'reading', u'books', u'materials', u'irish', u'would', u'led', u'author', u'critic', u'support', u'develo', u'greatly', u'enanced', u'ranging', u'teaching', u'materials', u'including', u'reading', u'irish', u'available', u'irish', u'me', u'range', u'reading', u'material', u'including', u'big', u'books', u'books', u'shared', u'reading', u'paired', u'reading', u'class', u'ls', u'feature', u'gaeltacht', u'speakers', u'des', u'continue', u'support', u'cogg', u'extending', u'teaching', u'materials', u'av', 'successfully connected to:127.0.0.1 27017 policy_test policy_test_extracted_cleaned
Public Policy Cleaned and pushed to mongodb Successfully!
```

Figure 39 Running Public Policy Data Extraction and Cleansing Module “CMD commands and output”



Key	Value	Type
_id	ObjectId("598b1a2db5803f2c6b3e4343")	Object
policy_row_text	120-year strategy for the irish language 2010...	String
policy_topic_terms	[0 elements]	Array
policy_cleaned_as_list_by_lines_or_pages	[46 elements]	Array
policy_name	policy_space	String

Figure 40 Public Policy Data Extraction and Cleansing Module Sample Output in mongo dB browser “Public Policy Data Structure in Mongodb”

¹² <https://github.com/mohadelrezk/GOV-DAF>

```

{
  "_id" : ObjectId("598b1a2db5803f2c6b3e4343"),
  "policy_row_text" : " 20-year strategy for the irish language 2010-2030: department of education and skills\n\n 2 20-year strategy for the irish language 2010-2030 the vision the objective of government policy in relation to irish is to increase on an incremental basis the use and knowledge of irish as a community language. specifically, the government's aim is to ensure that as many citizens as possible are bilingual in both irish and english. it is an integral component of the government's irish language policy that close attention be given to its place in the gaeltacht, particularly in light of research which indicates that the language's viability as a household and community language in the gaeltacht is under threat. the aim of government policy is also to: increase the number of families throughout the country who use irish as the daily language of communication; provide linguistic support for the gaeltacht as an irish-speaking community and to recognise the issues which arise in areas where irish is the household and community language; ensure that in public discourse and in public services the use of irish or english will be, as far as practical, a choice for the citizen to make and that over time more and more people throughout the state will choose to do their business in irish; and ensure that irish becomes more visible in our society, both as a spoken language by our citizens and also in areas such as signage and literature. irish is part of the culture and heritage of northern ireland and the promotion and protection of the language there is also a priority for the government. the government recognises the tremendous advantage to its citizens of fluency in english, the most widely used language in international affairs. the government commits to ensuring that this advantage is retained through the development of a bilingual society, where as many people as possible can use irish and english with equal ease and facility. however, the government also recognises that the focus of policy in the gaeltacht needs to be on maintaining the linguistic identity of the community in the gaeltacht as a distinctive language region, rather than one of bilingualism. while strengthening the position of the language within our education system is a key focus of this strategy, the transmission of irish as a living language within the family and between the generations is critically important. our overall approach is to create a supportive framework and the opportunities in which irish can be passed on in a natural way within households and communities. this is of special importance in the context of the gaeltacht.\n\n 3the future of the language depends on people who make a positive choice to embrace the opportunities that this"
}

```

Figure 41 Public Policy Data Extraction and Cleansing Module Sample Output in mongo dB browser “Public Policy as Text”

```

{
  "policy_cleaned_as_list_by_lines_or_pages" : [
    [
      "20",
      "year",
      "strategy",
      "irish",
      "language",
      "2010",
      "2030",
      "department",
      "education",
      "skills"
    ],
    [
      "20",
      "year",
      "strategy",
      "irish",
      "language",
      "2010",
      "2030",
      "vision",
      "objective",
      "government",
      "policy",
      "relation",
      "irish",
      "increase",
      "incremental",
      "basis",
      "use",
      "knowledge",
      "irish",
      "community",
      "language",
      "specifically",
      "government",
      "aim",
      "ensure",
      "many",
      "citizens",
      "possible",
      "bilingual",
      "irish",
      "culture"
    ]
  ]
}

```

Figure 42 Public Policy Data Extraction and Cleansing Module Sample Output in mongo dB browser “Public Policy as Word Vectors”

4.3.4 Public Policy Topic Modelling Module

```

GOV-DAF -- bash -- 198x10
/data — mongod + sudo ... ...LBACKUP/TEMP-WAITING-ORGANIZATION/3-workspace/GOV-DAF -- bash ...LBACKUP/TEMP-WAITI
WARNING:gensim.models.ldamodel:too few updates, training might not converge; consider increasing the number of passes or iterations to improve accuracy
('topic_terms': [('term': 'irish', 'score': u'0.031'), ('term': 'education', 'score': u'0.018'), ('term': 'language', 'score': u'0.018'), ('term': 'pri...re: u'0.013'), ('term': 'teaching', 'score': u'0.013'), ('term': 'medium', 'score': u'0.011'), ('term': 'gaeltacht', 'score': u'0.008'), ('term': 'sch...re: u'0.007')], 'policy_name': 'policy_space', 'topic_id': 0)
('topic_terms': [('term': 'irish', 'score': u'0.037'), ('term': 'language', 'score': u'0.019'), ('term': 'gaeltacht', 'score': u'0.014'), ('term': 'edu...core': u'0.011'), ('term': 'medium', 'score': u'0.011'), ('term': 'schools', 'score': u'0.010'), ('term': 'teaching', 'score': u'0.009'), ('term': 'pri...core': u'0.006')], 'policy_name': 'policy_space', 'topic_id': 1)
successfully connected to:127.0.0.1 27017 policy_test policy_test_extracted_cleaned_topicmodeled
Public Policy Topics detected and pushed to mongodb Successfully!

```

Figure 43 Running Public Policy Topic Modelling Module “CMD commands and output”

The screenshot shows the Robo 3T MongoDB browser interface. On the left, the database structure is visible, including collections like local, System, policy_test, and policy_test_extracted_cleaned_topicmodeled. The right panel displays the contents of the 'policy_test_extracted_cleaned_topicmodeled' collection. A specific document is selected, showing its fields and values. The 'topic_terms' field is expanded to show two arrays of terms and their scores. The first array (topic_id 0) contains terms like 'education' (score 0.018), 'language' (score 0.018), 'medium' (score 0.011), 'schools' (score 0.008), and 'teaching' (score 0.009). The second array (topic_id 1) contains terms like 'irish' (score 0.037), 'language' (score 0.019), 'gaeltacht' (score 0.014), 'medium' (score 0.011), and 'schools' (score 0.010).

Key	Value	Type
_id	Objectid("598b2716b5803f2dbe7e5e1b")	ObjectID
policy_row_text	120-year strategy for the Irish language 2...	String
topic_terms	[2 elements]	Array
topic_terms[0]	{ 3 fields }	Object
topic_terms[0].topic_terms	[10 elements]	Array
topic_terms[0].topic_terms[0]	{ 2 fields }	Object
topic_terms[0].topic_terms[1]	{ 2 fields }	Object
topic_terms[0].topic_terms[2]	{ 2 fields }	Object
topic_terms[0].topic_terms[3]	{ 2 fields }	Object
topic_terms[0].topic_terms[4]	{ 2 fields }	Object
topic_terms[0].topic_terms[5]	{ 2 fields }	Object
topic_terms[0].topic_terms[6]	{ 2 fields }	Object
topic_terms[0].topic_terms[7]	{ 2 fields }	Object
topic_terms[0].topic_terms[8]	{ 2 fields }	Object
topic_terms[0].topic_terms[9]	{ 2 fields }	Object
topic_terms[0].policy_name	policy_space	String
topic_terms[0].topic_id	0	Int32
topic_terms[1]	{ 3 fields }	Object
topic_terms[1].topic_terms	[10 elements]	Array
topic_terms[1].topic_terms[0]	{ 2 fields }	Object
topic_terms[1].topic_terms[1].term	irish	String
topic_terms[1].topic_terms[1].score	0.037	String
topic_terms[1].topic_terms[2]	{ 2 fields }	Object
topic_terms[1].topic_terms[3]	{ 2 fields }	Object
topic_terms[1].topic_terms[4]	{ 2 fields }	Object
topic_terms[1].topic_terms[5]	{ 2 fields }	Object
topic_terms[1].topic_terms[6]	{ 2 fields }	Object
topic_terms[1].topic_terms[7]	{ 2 fields }	Object
topic_terms[1].topic_terms[8]	{ 2 fields }	Object
topic_terms[1].topic_terms[9]	{ 2 fields }	Object
topic_terms[1].policy_name	policy_space	String
topic_terms[1].topic_id	1	Int32
topic_cleaned_as_list_by_lines_o...	[46 elements]	Array
topic_name	policy_space	String

Figure 44 Public Policy Topic Modelling Module Sample Output in mongo DB Browser “Public Policy Topics Vectors [2 topics]”

```

{
  "topic_terms": [
    {
      "term": "irish",
      "score": "0.037"
    },
    {
      "term": "language",
      "score": "0.019"
    },
    {
      "term": "gaeltacht",
      "score": "0.014"
    },
    {
      "term": "education",
      "score": "0.014"
    },
    {
      "term": "des",
      "score": "0.011"
    },
    {
      "term": "medium",
      "score": "0.011"
    },
    {
      "term": "schools",
      "score": "0.010"
    },
    {
      "term": "teaching",
      "score": "0.009"
    },
    {
      "term": "primary",
      "score": "0.008"
    },
    {
      "term": "students",
      "score": "0.006"
    }
  ],
  "policy_name": "policy_space",
  "topic_id": 1
}

```

Figure 45 Public Policy Topic Modelling Module Sample Output in mongo DB Browser “Public Policy Topic Vector”

4.3.5 Public Policy Co-occurrence Matrix Building Module

```

[<CitizenSatisfactionAnalysis.Correlation_cell instance at 0x1130cce18>, <CitizenSatisfactionAnalysis.Correlation_cell instance at 0x1130cce18>, <CitizenSatisfactionAnalysis.Correlation_cell instance at 0x1130cce9e0>, <CitizenSatisfactionAnalysis.Correlation_cell instance at 0x1130cc248>, <CitizenSatisfactionAnalysis.Correlation_cell instance at 0x1130cc290>, <CitizenSatisfactionAnalysis.Correlation_cell instance at 0x1130cc950>, <CitizenSatisfactionAnalysis.Correlation_cell instance at 0x1130cc0e0>, <CitizenSatisfactionAnalysis.Correlation_cell instance at 0x1130ccdd0>, <CitizenSatisfactionAnalysis.Correlation_cell instance at 0x1130ccf0>, <CitizenSatisfactionAnalysis.Correlation_cell instance at 0x1130cc320>, <CitizenSatisfactionAnalysis.Correlation_cell instance at 0x1130cc758>, <CitizenSatisfactionAnalysis.Correlation_cell instance at 0x1130cc908>, <CitizenSatisfactionAnalysis.Correlation_cell instance at 0x1130cc6c8>, <CitizenSatisfactionAnalysis.Correlation_cell instance at 0x1130cc1b8>, <CitizenSatisfactionAnalysis.Correlation_cell instance at 0x1130cc0c0>, <CitizenSatisfactionAnalysis.Correlation_cell instance at 0x1130ce170>, <CitizenSatisfactionAnalysis.Correlation_cell instance at 0x1130ce440>, <CitizenSatisfactionAnalysis.Correlation_cell instance at 0x1130ce560>, <CitizenSatisfactionAnalysis.Correlation_cell instance at 0x1130ce5f0>, <CitizenSatisfactionAnalysis.Correlation_cell instance at 0x1130cel28>, <CitizenSatisfactionAnalysis.Correlation_cell instance at 0x1130cea70>, <CitizenSatisfactionAnalysis.Correlation_cell instance at 0x1130ce950>, <CitizenSatisfactionAnalysis.Correlation_cell instance at 0x1130ce9e0>, <CitizenSatisfactionAnalysis.Correlation_cell instance at 0x1130ceb00>, <CitizenSatisfactionAnalysis.Correlation_cell instance at 0x1130ceb90>, <CitizenSatisfactionAnalysis.Correlation_cell instance at 0x1130ce050>]
successfully connected to:127.0.0.1 27017 policy_test policy_test_correlationMatrix
Public Policy co-occurrences matrix constructed and pushed to mongodb Successfully!

```

Figure 46 Running Public Policy Co-occurrence Matrix Building Module “CMD commands and output”

Figure 47 Public Policy Co-occurrence Matrix Building Module Sample Output in mongo DB Browser “Public Policy Co-occurrence Matrix”

4.3.6 Public Policy Related Tweets Collection Module

```
(venv) mohade-mackpro:GOV-DAF mohade$ python TweeterProcessing.py
read from config file!
connect
get tweets
successfully connected to:127.0.0.1    27017    test_policy    policy_analysis2
{
  "search_metadata": {
    "completed_in": 0.054,
    "count": 15,
    "max_id": "895321447861698562",
    "max_id_str": "895321447861698562",
    "next_results": "?max_id=895321425824800767&q=politic&include_entities=1",
    "query": "politic",
    "refresh_url": "?since_id=895321447861698562&q=politic&include_entities=1",
    "since_id": 0,
    "since_id_str": "0"
  },
  "statuses": [
    {
      "contributors": null,
      "coordinates": null,
      "created_at": "Wed Aug 09 16:30:42 +0000 2017",
      "entities": {
        "hashtags": [
          {
            "indices": [
              79,
              84
            ],
            "text": "silly"
          }
        ],
        "symbols": [],
        "urls": [
          {
            "display_url": "buff.ly/2wt86GG",
            "expanded_url": "https://buff.ly/2wt86GG",
            "indices": [
              55,
              78
            ],
            "url": "https://t.co/huz6qMs6"
          }
        ]
      },
      "source": "<a href=\"https://www.Arab-girls-for-dating.com\" rel=\"nofollow\">Arab girls for dating</a>",
      "text": "\u0634\u0627\u0647\u062f \u0645\u0627\u0630\u0627 \u0641\u0639\u0644 \u0634\u0627\u0628 \u0628\u0641\u062a\u0645\u0627\u0647\u0631\u0629.. https://t.co/yd8367AkPS\n\n\u0645\u0628\u0627\u0647\u0631\u0629 https://t.co/huz6qMs6",
      "truncated": false,
      "user": {
        "contributors_enabled": false,
        "created_at": "Fri Sep 30 20:55:19 +0000 2016",
        "default_profile": true,
        "default_profile_image": true,
        "description": "",
        "entities": {
          "description": {
            "urls": []
          }
        },
        "favourites_count": 0,
        "follow_request_sent": false,
        "followers_count": 14,
        "following": false,
        "friends_count": 21,
        "geo_enabled": false,
        "has_extended_profile": false,
        "id": 781960641544421376,
        "id_str": "781960641544421376",
        "is_translation_enabled": false,
        "is_translator": false,
        "lang": "ru",
        "listed_count": 0,
        "location": "",
        "name": "\u042e\u043b\u0438\u044f \u0430\u0431\u0437\u0430\u0430\u043b\u043e\u0432\u0430",
        "notifications": false,
        "profile_background_color": "F5F8FA",
        "profile_background_image_url": null,
        "profile_background_image_url_https": null,
        "profile_background_tile": false,
        "profile_image_url": "http://abs.twimg.com/sticky/default_profile_images/default_profile_normal.png",
        "profile_image_url_https": "https://abs.twimg.com/sticky/default_profile_images/default_profile_normal.png",
        "profile_link_color": "#1DA1F2",
        "profile_sidebar_border_color": "#C0DDE6",
        "profile_sidebar_fill_color": "#DDEEF6",
        "profile_text_color": "#333333",
        "profile_use_background_image": true,
        "protected": false,
        "screen_name": "MikheySukhanov",
        "statuses_count": 9896,
        "time_zone": "Pacific Time (US & Canada)",
        "translator_type": "none",
        "url": null,
        "utc_offset": -25200,
        "verified": false
      }
    }
  ]
}
Public Policy related tweets collected and pushed to mongodb Succesfully!
```

Figure 48 Running Public Policy Related Tweet Collection Module “CMD commands and output”

```

Robo 3T - 1.1

db.getCollection('policy_analysis2').find()
  local   localhost:27017  test_policy
  db.getCollection('policy_analysis2').find({})

policy_analysis2  0.002 sec.

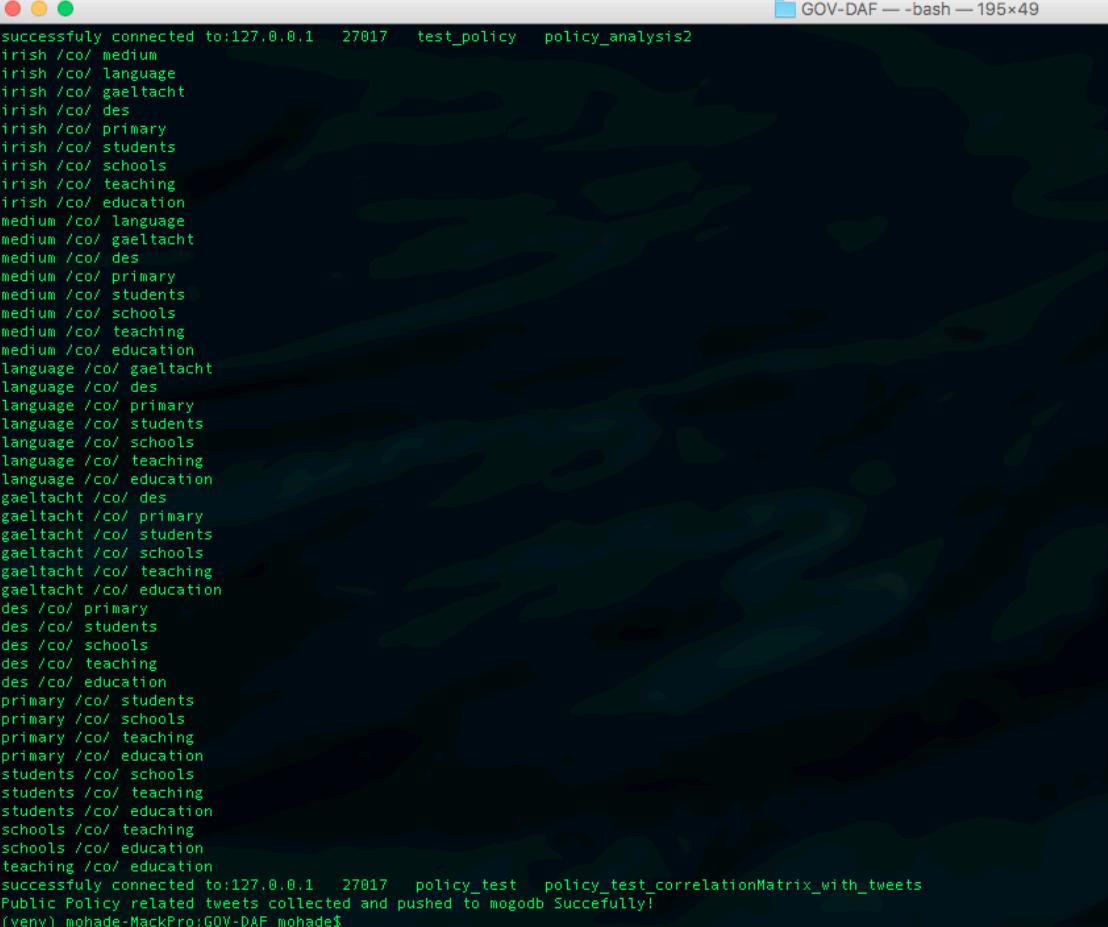
Key          Value
---          ---
(1) ObjectId("598b397fb5803f2ea5d42bda") { 29 fields }
(2) ObjectId("598b3980b5803f2ea5d42bd0") { 26 fields }
(3) ObjectId("598b3980b5803f2ea5d42bdc") { 28 fields }
(4) ObjectId("598b3980b5803f2ea5d42bdd") { 28 fields }
(5) ObjectId("598b3980b5803f2ea5d42bde") { 29 fields }
(6) ObjectId("598b3980b5803f2ea5d42bdf") { 27 fields }

_id           ObjectId("598b3980b5803f2ea5d42bd0")
contributors  null
truncated    false
text          RT @votevets: Air Force vet Capt. "Sully" Sullenberger joins the opposition against @POTUS's push ...
is_quote_status false
in_reply_to_status_id null
id            895322221471604736
favorite_count 0
entities      { 4 fields }
retweeted    false
coordinates  null
source        <a href="http://twitter.com/download/android" rel="nofollow">Twitter for Android</a>
in_reply_to_screen_name null
in_reply_to_user_id null
retweet_count 404
id_str        895322221471604736
favorited    false
retweeted_status { 25 fields }
user          { 42 fields }
geo           null
possibly_sensitive false
lang          en
created_at    Wed Aug 09 16:33:46 +0000 2017
in_reply_to_status_id_str null
place         null
metadata     { 2 fields }

```

Figure 49 Public Policy Related Tweet Collection Module Sample Output in mongo DB Browser “Public Policy Related Tweets”

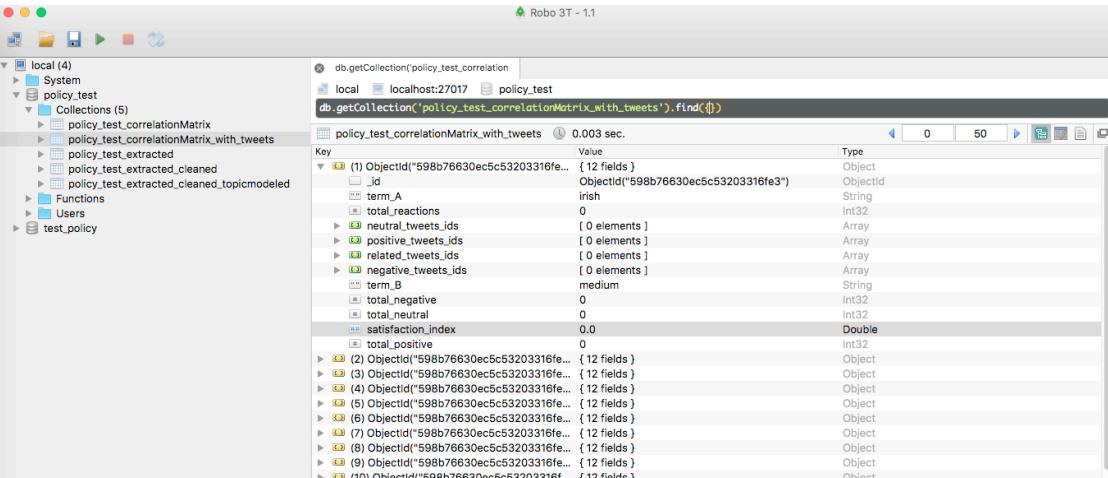
4.3.7 Public Policy Matrix Based Related Tweets Querying Module



```

successfully connected to:127.0.0.1 27017 test_policy policy_analysis2
irish /co/ medium
irish /co/ language
irish /co/ gaeltacht
irish /co/ des
irish /co/ primary
irish /co/ students
irish /co/ schools
irish /co/ teaching
irish /co/ education
medium /co/ language
medium /co/ gaeltacht
medium /co/ des
medium /co/ primary
medium /co/ students
medium /co/ schools
medium /co/ teaching
medium /co/ education
language /co/ gaeltacht
language /co/ des
language /co/ primary
language /co/ students
language /co/ schools
language /co/ teaching
language /co/ education
gaeltacht /co/ des
gaeltacht /co/ primary
gaeltacht /co/ students
gaeltacht /co/ schools
gaeltacht /co/ teaching
gaeltacht /co/ education
des /co/ primary
des /co/ students
des /co/ schools
des /co/ teaching
des /co/ education
primary /co/ students
primary /co/ schools
primary /co/ teaching
primary /co/ education
students /co/ schools
students /co/ teaching
students /co/ education
schools /co/ teaching
schools /co/ education
teaching /co/ education
successfully connected to:127.0.0.1 27017 policy_test policy_test_correlationMatrix_with_tweets
Public Policy related tweets collected and pushed to mongodb Succesfully!
(venv) mohade-MackPro:GOV-DAF mohade$
```

Figure 50 Running Public Policy Matrix Based Related Tweets Querying Module “CMD commands and output”



Key	Value	Type
term_A	irish	String
total_reactions	0	Int32
neutral_tweets_ids	[]	Array
positive_tweets_ids	[]	Array
related_tweets_ids	[]	Array
negative_tweets_ids	[]	Array
term_B	medium	String
total_negative	0	Int32
total_neutral	0	Int32
satisfaction_index	0.0	Double
total_positive	0	Int32

Figure 51 Public Policy Matrix Based Related Tweets Querying Module Sample Output in mongo DB Browser “Public Policy Co-occurrence cell Related Tweets”

```

{
    "_id" : ObjectId("598b76630ec5c53203316fe3"),
    "term_A" : "irish",
    "total_reactions" : 0,
    "neutral_tweets_ids" : [],
    "positive_tweets_ids" : [],
    "related_tweets_ids" : [],
    "negative_tweets_ids" : [],
    "term_B" : "medium",
    "total_negative" : 0,
    "total_neutral" : 0,
    "satisfaction_index" : 0.0,
    "total_positive" : 0
}

```

Figure 52 Public Policy Matrix Based Related Tweets Querying Module Sample Output in mongo DB Browser “Public Policy Co-occurrence cell Related Tweets” Document view

4.3.8 Public Policy Sentiment Analysis Module

Key	Value	Type
(1) ObjectId("598b76630ec5c53203316fe3")	{ 12 fields }	Object
_id	ObjectId("598b76630ec5c53203316fe3")	ObjectID
term_A	irish	String
total_reactions	14	Int32
neutral_tweets_ids	[4 elements]	Array
positive_tweets_ids	[2 elements]	Array
related_tweets_ids	[14 elements]	Array
negative_tweets_ids	[8 elements]	Array
term_B	medium	String
total_negative	8	Int32
total_neutral	4	Int32
satisfaction_index	0.14285714285714285	Double
total_positive	2	Int32

Figure 53 Public Policy Sentiment Analysis Module Sample Output in mongo DB Browser

The screenshot shows a mongo DB Browser window with the title "View Document". The address bar indicates the connection is to "localhost:27017" and the database/collection is "policy_test / policy_test_correlationMatrix_with_tweets". The main content area displays a single document as a JSON object. The document contains various fields such as "_id", "term_A", "total_reactions", "neutral_tweets_ids", "positive_tweets_ids", "related_tweets_ids", "negative_tweets_ids", "term_B", and several numerical and array fields. The "negative_tweets_ids" field is particularly large, listing numerous tweet IDs. A "Cancel" button is visible at the bottom right of the document view.

```
{  
    "_id" : ObjectId("598b76630ec5c53203316fe3"),  
    "term_A" : "irish",  
    "total_reactions" : 14,  
    "neutral_tweets_ids" : [  
        NumberLong(895322244829782016),  
        NumberLong(895322244829782014),  
        NumberLong(895322244829790000),  
        NumberLong(895322244829782890)  
    ],  
    "positive_tweets_ids" : [  
        NumberLong(895322244829782016),  
        NumberLong(895322244829782999)  
    ],  
    "related_tweets_ids" : [  
        NumberLong(895322244829782016),  
        NumberLong(895322244829782014),  
        NumberLong(895322244829790000),  
        NumberLong(895322244829782890),  
        NumberLong(895322244829782016),  
        NumberLong(89532224482978222),  
        NumberLong(895322244829789012),  
        NumberLong(895322244829782054),  
        NumberLong(895322244829782000),  
        NumberLong(895322244829782016),  
        NumberLong(89532224482978277),  
        NumberLong(895322244829782099),  
        NumberLong(895322244829782016),  
        NumberLong(895322244829782999)  
    ],  
    "negative_tweets_ids" : [  
        NumberLong(895322244829782016),  
        NumberLong(89532224482978222),  
        NumberLong(895322244829789012),  
        NumberLong(895322244829782054),  
        NumberLong(895322244829782000),  
        NumberLong(895322244829782016),  
        NumberLong(89532224482978277),  
        NumberLong(895322244829782099)  
    ],  
    "term_B" : "medium",  
    "total_negative" : 8,  
    "total_neutral" : 4,  
    "satisfaction_index" : 0.14285714285,  
    "total_positive" : 2  
}
```

Figure 54 Public Policy Sentiment Analysis Module Sample Output in mongo DB Browser “Document view”

4.3.9 Building Prediction Model and Citizen Satisfaction Estimation Module

Using the following Formulas over the output of last phases:

Equation 3 Citizen Satisfaction Analysis Functions

$$\begin{aligned} topicspace &= \int (policy space) \\ topic(policy'a') &= \{term'1', term'2' ... term'n'\} \\ tweets(policy'a') &= \{tweet'1-n' | contains two terms from terms(policy'a')\} \\ \text{actual sentiment profile } (policy'a') &= \sum \text{deep learning sentiments}(tweets(policy'a')) \\ \text{predicted sentiment profile}(policy'p') &= \sum \frac{p(\text{term}'x' \cap \text{term}'y' | \text{negative} | \text{positive} | \text{neutral})}{p(\text{term}'x' \cap \text{term}'y')} \end{aligned}$$

..where :

- [1] policy 'a' is a current public policy
- [2] policy 'p' is a public policy in progress
- [3] prediction applied by Bayes Theorem
- [4] topics extracted using LDA Theroy
- [5] sentiments calculated using deep learning algorithm
- [6] p is probability
- [7] sentiment profile is (total negative, total positive and total neutral tweets)

4.4 Conclusion

In this chapter “Experiments and Evaluation” we demonstrated the Gov-DAF components implementation details, in the next chapter “Conclusion” we will discuss research results and open issues.

5 CONCLUSION

In this chapter we will discuss the outcomes and issues of the different research components. Following Experimentation, Evaluation and Limitation and Open Research Issues of the research.

5.1 *Experimentation, Evaluation and Limitation*

Following Experimentation, Evaluation and Limitation discussion across research components:

5.1.1 Core Public Policy Vocabulary

We highlight application scenarios for the CPPV vocabulary.

Scenario 1: Semantic Policy Search Engine: Researchers and Business users of a CPPV ontology based search engine to browse the public policies or run complex queries, to discover opportunities in a market. Example: Firm X is specialized in the energy sector; they would like to extend their business to Country Y. To get a better understanding of the market they need to review Country Y's energy policies and related policies using a search engine such as the semantic policy search engine.

Scenario 2: "Policy learning" - During the policy making or policy update processes, governments may review policies from the Open Public Policy platform to find best practices or certain policy aspects to adopt. Example: Government Y is developing a new Health policy and would want to review different Health policies implemented in the European Union, query for certain aspects of the Health policies to improve their policy decisions.

5.1.2 Formal Adoption of the Core Public Policy Vocabulary

We briefly highlight the vision for the adoption of the vocabulary. Currently, the vocabulary is being reviewed by experts involved in public sector information modelling practices in Ireland, Scotland and UK, semantic interoperability and

standards in Europe for staged adoption. After the adoption by a these governments, the follow-up steps entails submitting the vocabulary to the European Commission's ISA programme as a candidate Core Vocabulary.

Field	Value
Policy Level	National
Policy Country	Ireland
Policy Language	English
Policy Start Date	2015-01-01
Policy Document Name	Energy White Paper - Dec 2015.pdf
Policy Document Description	Energy White Paper - Dec 2015.pdf
Policy Document File Type	pdf
Policy Document URL	http://www.dccae.gov.ie/energy/Lists/Publications%20Documents/Energy%20White%20Paper%20-%20Dec%202015.pdf
Vision(s)	To transform Ireland into a low carbon society and economy by 2050

Figure 55 Core Public Policy Vocabulary Implementations in Protégé and CKAN

5.1.3 Public Policy Collaboration Mining for Citizen Satisfaction

We highlight application areas for the Public Policy Collaboration Mining.

Application Area 1: Standardization and Collaboration Analysis: Despite most of governments already publishing their data via their open data portals, when a government decides to integrate their data sources over its variant departments and councils, this heterogeneous domain dependent data will consume huge analysis resources and a considerably extended period of time to be fitted into an integrated data repository. Our profiling service will lead the way for data analysts to define integration channels, and necessary concepts standardizations between governmental departments and councils, using the available data published on open data portals. Same example would fit a multinational enterprise as well.

For example “marine-institute” and “geological-survey-of-ireland” share the named entity (pollution), this concept shall be standardized regarding its code and its

measurement unit to ease integration and comparability or analysis I general among multiple datasets.

Application Area 2: Intelligent Open Data Portals Exploration: Open data portals are meant to be facing the public in other words the citizens, but citizens can't directly comprehend, and consume this raw data [4]. Open data portals profiling service will help citizens to easily and intelligently explore the open data portal using visualized publishers and dataset profiles.

Finally, Named Entity Recognition area of the work is tightly coupled with the training and the quality of the Named Entity Recognition algorithm. Through this research we have experimented Natural Language Tool Kit (NLTK), Stanford NER and Stanford NER with nGram of (3) enhancement, then we ended up using DBpedia Spotlight as the NE source as through our manual examination of the text analysis phase results DBpedia out performed the other methods in its NE detection quality. DBpedia spotlight still have its limitations though and we reported one of the issues we faced to their github repository¹³.

¹³ <https://github.com/dbpedia-spotlight/dbpedia-spotlight/issues/407>

The figure displays two side-by-side screenshots of the Data.gov.ie website. Both screens show search results for the term "pollution".

Left Screen (Marine Institute Datasets):

- Search Bar:** pollution
- Sort By:** Relevance
- Results Count:** 7 Results
- Dataset 1:** Irish Groundfish Survey Catch Weights (Environment, 3 stars)
- Dataset 2:** Irish Groundfish Survey Catch Numbers (Environment, 3 stars)
- Dataset 3:** Irish Groundfish Survey Stations (Environment, 3 stars)
- Dataset 4:** Contaminants in water in the marine environment (Environment, 3 stars)
- Dataset 5:** Contaminants in sediments in the marine environment (Environment, 3 stars)
- Dataset 6:** Burrishole Catchment Automatic River Monitoring Stations [ARMS] (Environment, 3 stars)
- Dataset 7:** Contaminants in biota in the marine environment (Environment, 3 stars)

Right Screen (Geological Survey of Ireland Datasets):

- Search Bar:** pollution
- Sort By:** Relevance
- Results Count:** 7 Results
- Dataset 1:** GSI Source Protection Areas (Science and Technology, 3 stars)
- Dataset 2:** GSI Groundwater Reports (Science and Technology, 3 stars)
- Dataset 3:** GSI Groundwater Tracer Lines (Science and Technology, 3 stars)
- Dataset 4:** GSI Groundwater Karst Features (Science and Technology, 3 stars)
- Dataset 5:** GSI Groundwater Vulnerability (Science and Technology, 3 stars)
- Dataset 6:** GSI Groundwater Sand and Gravel Aquifers (Science and Technology, 3 stars)
- Dataset 7:** GSI Groundwater Recharge Mapping (Science and Technology, 3 stars)

Figure 56 Datasets shared between Marine Institute and Geological Survey of Ireland around the concept pollution¹⁴

¹⁴ <https://data.gov.ie/data/search?q=pollution&publisher=marine-institute>

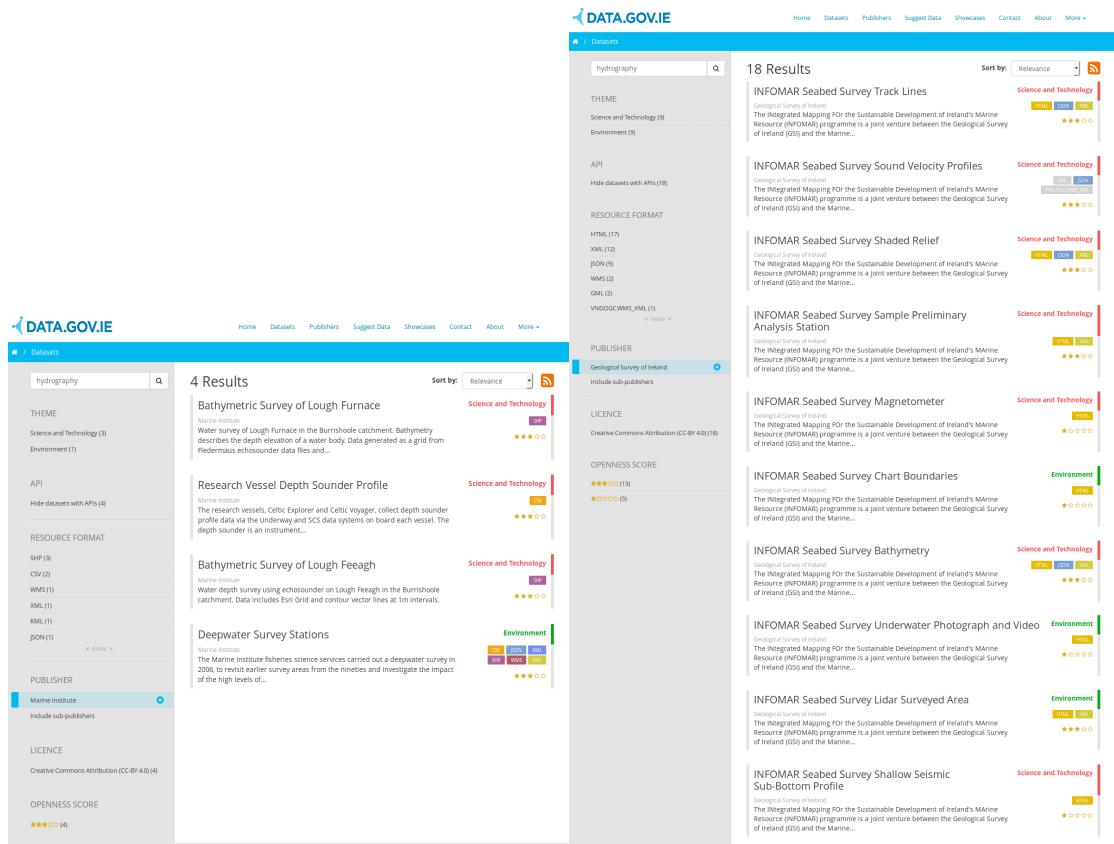


Figure 57 Datasets shared between Marine Institute and Geological Survey of Ireland around the concept hydrography^{15 16}

5.1.4 Gov – DAF

As demonstrated in chapters 3 and 4 Gov-DAF component of the research is used to measure the satisfaction of the citizen towards a public policy. This can be done either for a current policy or an in progress policy by implementing to the Gov-DAF pipeline Figure 1 and formulas Equation 1 and Equation 3.

Finally, Gov-DAF results accuracy are tightly coupled with tow factors, A. Tweet dataset size and quality, quality means its relatedness with the public policy in discussion. B. Prediction algorithm applied on the matrix to estimate citizen satisfaction for an in progress public policy.

¹⁵ <https://data.gov.ie/data/search?q=hydrography&publisher=marine-institute>

¹⁶ <https://data.gov.ie/data/search?q=hydrography&publisher=geological-survey-of-ireland>

5.2 Future Work and Open Research Issues

5.2.1 Core Public Policy Vocabulary

The CPPV was motivated by the lack of a central repository for government policies in Europe and the underpinning semantic data model or vocabulary to facilitate the inter-linking and exchange of public policies in the European Public Administration Ecosystem. The unavailability of such infrastructure makes the collaboration among public agencies and development of coherent public policies difficult. In addressing this challenge, we have developed a public policy ontology guided by the principles driving the ISA's Core Vocabulary initiatives. Our Core Public Policy Vocabulary complements existing core vocabularies and reuses elements of these vocabularies.

We are currently in the second stage of the third-party country expert review process of the vocabulary. We intend to develop many use cases to demonstrate policy collaboration opportunities afforded by the CPPV. In addition, we plan to provide reference implementation using other technologies such as Semantic Media Wiki or Wiki Data¹⁷. In the near future, we intend to propose the CPPV vocab as a candidate ISA Core Vocabulary.

5.2.2 Public Policy Collaboration Mining for Citizen Satisfaction

Regarding our approach results we believe that we are on the right track to tackle the collaboration mining problem in open governmental data domain, as we are getting interested collaboration recommendations out of our pipeline in a visualized way that is easy to comprehend by general public users of open governmental data.

Our future plan is to overcome the NE limitation by developing a new text analysis pipeline that integrates statistical text analysis, babel.net¹⁸, and DBpedia¹⁹ as our NE source. Also we are planning to replace the string comparison module with semantic relatedness comparison module as the way of calculating relation strength between open governmental data publishers.

¹⁷ https://www.wikidata.org/wiki/Wikidata:Main_Page

¹⁸ <http://babelnet.org/>

¹⁹ <http://wiki.dbpedia.org/>

5.2.3 Gov-DAF

Gov-DAF was proposed in [11], [12], Gov-DAF is addressing the problem of lack of tools to support critical government decision making in which knowledge of citizen opinions expressed on social media constitute a critical input. Now we are presenting Gov-DAF knowledge base component known as Gov-DAF knowledge base building pipeline implementation in details, Gov-DAF Knowledge base building pipeline was implemented using two methods Named Entity Recognition Based Methodology, and Topic Modelling Based Methodology. Both Methods implementation details are reported and accuracy indices have been measured highlighted and presented. Reckoning to the accuracy indices results Gov-DAF will adopt and enhance Topic Modelling Based Methodology for the next phases of Gov-DAF implementation namely (Opinion Harvesting, Opinion Mining and Satisfaction Estimation).

BIBLIOGRAPHY

- [1] S. de Walle and G. G. Van Ryzin, “The Order Of Questions In A Survey On Citizen Satisfaction With Public Services: Lessons From A Split-ballot Experiment,” *Public Adm.*, vol. 89, no. 4, pp. 1436–1450, 2011.
- [2] J. M. Kelly, “Citizen Satisfaction and Administrative Performance Measures Is there Really a Link?,” *Urban Aff. Rev.*, vol. 38, no. 6, pp. 855–866, 2003.
- [3] J. M. Kelly and D. Swindell, “A multiple--indicator approach to municipal service evaluation: correlating performance measurement and citizen satisfaction across jurisdictions,” *Public Adm. Rev.*, vol. 62, no. 5, pp. 610–621, 2002.
- [4] S. L. Percy, “Response time and citizen evaluation of police,” *J. Police Sci. Adm.*, vol. 8, no. 1, pp. 75–86, 1980.
- [5] G. G. Ryzin, D. Muzzio, S. Immerwahr, L. Gulick, and E. Martinez, “Drivers and consequences of citizen satisfaction: An application of the American customer satisfaction index model to New York City,” *Public Adm. Rev.*, vol. 64, no. 3, pp. 331–341, 2004.
- [6] W. G. Skogan, “Citizen satisfaction with police encounters,” *Police Q.*, vol. 8, no. 3, pp. 298–321, 2005.
- [7] B. Stipak, “Citizen satisfaction with urban services: Potential misuse as a performance indicator,” *Public Adm. Rev.*, pp. 46–52, 1979.
- [8] D. Swindell and J. M. Kelly, “Linking citizen satisfaction data to performance measures: A preliminary evaluation,” *Public Perform. Manag. Rev.*, pp. 30–52, 2000.
- [9] G. G. Van Ryzin, “Expectations, performance, and citizen satisfaction with urban services,” *J. Policy Anal. Manag.*, vol. 23, no. 3, pp. 433–448, 2004.
- [10] E. W. Welch, C. C. Hinnant, and M. J. Moon, “Linking citizen satisfaction with e-government and trust in government,” *J. public Adm. Res. theory*, vol. 15, no. 3, pp. 371–391, 2005.
- [11] M. Adel Rezk, A. Ojo, G. A. El Khayat, and S. Hussein, “A Proposed Government Decision Support System Based on Citizens Interactions over Social Networks,” in *Proceedings of the FIFTH International Conference on Information and Communication Technology in Our Lives 2015*, 2015.

- [12] M. A. Rezk, A. Ojo, G. A. El Khayat, and S. Hussein, “A Government Decision Analytics Framework Based on Citizen Opinion,” in *9th International Conference on Theory and Practice of Electronic Governance*, 2016.
- [13] M. Adel Rezk, M. H. Aliyu, H. Bensta, and A. Ojo, “Core Public Policy Vocabulary.”
- [14] M. Adel Rezk, A. Ojo, G. El Khayat, and S. Hussien, “A Government Decision Analytics Framework Based on Citizen Opinion (Gov-DAF): Elaboration of the Knowledge Base Component,” in *6th International Conference on Information and Communication Technology*, 2016.
- [15] M. A. Rezk, M. H. Aliyu, H. Bensta, and A. Ojo, “Standardising Public Policy Documentation to foster Collaboration across Government Agencies 1 Introduction,” pp. 1–12.
- [16] M. A. Rezk, M. H. Aliyu, H. Bensta, and A. Ojo, “Towards a Vocabulary for Open Public Policies,” in *Proceedings of the 10th International Conference on Theory and Practice of Electronic Governance - ICEGOV '17*, 2017, pp. 584–585.
- [17] R. Bandari, S. Asur, and B. A. Huberman, “The Pulse of News in Social Media: Forecasting Popularity,” *arXiv Prepr. arXiv1202.0332*, 2012.
- [18] Peter Teufl, Udo Payer, and Peter Parycek, “Automated Analysis of e-Participation Data by Utilizing Associative Networks, Spreading Activation and Unsupervised Learning.”
- [19] N. Shadbolt *et al.*, “Linked open government data: Lessons from data.gov.uk,” *IEEE Intell. Syst.*, vol. 27, no. 3, pp. 16–24, 2012.
- [20] K. Breitman, P. Salas, M. A. Casanova, and D. Saraiva, “Open Government Data in Brazil,” *Intell. Syst. IEEE*, vol. 27, no. 3 (May-June), pp. 45–49, 2012.
- [21] L. N. Mutuku and J. Colaco, “Increasing Kenyan open data consumption,” *Proc. 6th Int. Conf. Theory Pract. Electron. Gov. - ICEGOV '12*, p. 18, 2012.
- [22] F. Artigas and S. A. Chun, “Visual analytics for open government data,” *14th Annu. Int. Digit. Gov. Res. Conf. From E-Government to Smart Gov. dg.o 2013*, pp. 298–299, 2013.
- [23] D. C. Ribeiro, J. Freire, H. T. Vo, and C. T. Silva, “An Urban Data Profiler,” *WWW Work. Data Sci. Smart Cities*, pp. 1389–1394, 2015.
- [24] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives,

- Dbpedia: A nucleus for a web of open data.* Springer, 2007.
- [25] P. N. Mendes, M. Jakob, A. García-Silva, and C. Bizer, “DBpedia spotlight: shedding light on the web of documents,” in *Proceedings of the 7th International Conference on Semantic Systems*, 2011, pp. 1–8.
 - [26] R. Sharda, S. H. Barr, and J. C. McDonnell, “Decision Support System Effectiveness: A Review and an Empirical Test,” *Manage. Sci.*, vol. 34, no. 2, pp. 139–159, Feb. 1988.
 - [27] R. H. Sprague and E. D. Carlson, *Building effective decision support systems*. Prentice-Hall, 1982.
 - [28] D. J. Power, “Decision Support Systems: A Historical Overview,” in *Handbook on Decision Support Systems 1*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 121–140.
 - [29] S. Alter, *Decision support systems : current practice and continuing challenges*. Addison-Wesley Pub, 1980.
 - [30] M. Alavi and D. Leidner, “Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues,” *Manag. Inf. Syst. Q.*, vol. 25, no. 1, Mar. 2001.
 - [31] J. R. Finkel, T. Grenager, and C. Manning, “Incorporating non-local information into information extraction systems by gibbs sampling,” in *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, 2005, pp. 363–370.
 - [32] K. J. Rodriguez, M. Bryant, T. Blanke, and M. Luszczynska, “Comparison of Named Entity Recognition tools for raw OCR text,” 2012.
 - [33] A. Esuli and F. Sebastiani, “SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining.”
 - [34] Umadevi V, “Sentiment Analysis Using Weka,” *Int. J. Eng. Trends Technol.*, vol. 18, no. 4, 2014.
 - [35] R. Prabowo and M. Thelwall, “Sentiment Analysis: A Combined Approach.”
 - [36] P. Kolb, “Disco: A multilingual database of distributionally similar words,” *Proc. KONVENS-2008, Berlin*, 2008.
 - [37] P. Kolb, “Experiments on the difference between semantic similarity and relatedness,” in *Proceedings of the 17th Nordic Conference on Computational Linguistics-NODALIDA’09*, 2009.

- [38] M. Beseiso, A. Ahmad, and R. Ismail, “A survey of Arabic language Support in Semantic web.”
- [39] K. Shaalan, “A Survey of Arabic Named Entity Recognition and Classification.”
- [40] A. al Owisheq, S. al Humoud, N. al Twairesh, and T. al Buhairi, “Arabic sentiment analysis resources: A survey,” 2016, vol. 9742, pp. 267–278.
- [41] M. Korayem, D. Crandall, and M. Abdul-Mageed, “Subjectivity and Sentiment Analysis of Arabic: A Survey,” 2012, vol. 322, pp. 128–139.
- [42] K. Almeman and M. Lee, “Automatic building of Arabic multi dialect text corpora by bootstrapping dialect words,” 2013.
- [43] A. M. Al-Thubaity, M. Khan, S. Alotaibi, and B. Alonazi, “Automatic Arabic term extraction from special domain corpora,” 2014, pp. 1–5.
- [44] R. B. Bahloul, M. Elkarwi, K. Haddar, and P. Blache, “Building an Arabic linguistic resource from a treebank: The case of property grammar,” 2014, vol. 8655 LNAI, pp. 240–246.
- [45] A. M. Azmi and R. S. Almajed, “A survey of automatic Arabic diacritization techniques,” vol. 21, no. 3, pp. 477–495, May 2015.
- [46] M. Belgacem and M. Zrigui, “Automatic identification system of arabic dialects,” 2010, vol. 2, pp. 740–749.
- [47] A. Alqudsi, N. Omar, and K. Shaker, “Arabic machine translation: a survey,” vol. 42, no. 4, pp. 549–572, 2012.
- [48] M. Biltawi, W. Etaifi, S. Tedmori, A. Hudaib, and A. Awajan, “Sentiment classification techniques for Arabic language: A survey,” 2016, pp. 339–346.
- [49] N. Al-Twairesh, H. Al-Khalifa, and A. Al-Salman, “Subjectivity and sentiment analysis of Arabic: Trends and challenges,” 2014, vol. 2014, pp. 148–155.
- [50] A. Hamdi, K. Shaban, and A. Zainal, “A Review on Challenging Issues in Arabic Sentiment Analysis,” *J. Comput. Sci.*, vol. 12, no. 9, pp. 471–481, Sep. 2016.
- [51] A. B. Al-Saleh and M. E. B. Menai, “Automatic Arabic text summarization: a survey,” vol. 45, no. 2, pp. 203–234, Feb. 2016.
- [52] A. H. Ababneh, J. Lu, and Q. Xu, “Arabic information retrieval: A relevancy assessment survey,” 2016, pp. 345–357.
- [53] A. Dandashi, J. Al Jaam, and S. Foufou, “Arabic named entity recognition—a

- survey and analysis,” 2016, vol. 55, pp. 83–96.
- [54] K. Shaalan, “A Survey of Arabic Named Entity Recognition and Classification,” vol. 40, no. 2, pp. 469–510, 2014.
 - [55] D. Nadeau, “A survey of named entity recognition and classification,” *Linguist. Investig.*, no. 30, p. 3–26., 2007.
 - [56] M. Albared, N. Omar, and M. J. Ab Aziz, “Arabic part of speech disambiguation: A survey,” vol. 4, no. 5, pp. 517–532, Sep. 2009.
 - [57] I. A. Al-Sughaiyer and A. Al-Kharashi, “Arabic morphological analysis techniques: A comprehensive survey,” vol. 55, no. 3, pp. 189–213, Feb. 2004.
 - [58] S. Albrecht *et al.*, “eParticipation – Electronic Participation of Citizens and the Business Community in eGovernment",” 2008.
 - [59] E. Tambouris, E. Kalampokis, and K. Tarabanis, “A survey of e-participation research projects in the European Union,” *Int. J. Electron. Bus.*, vol. 6, no. 6, p. 554, 2008.
 - [60] E. Bonsón, S. Royo, and M. Ratkai, “Citizens’ engagement on local governments’ Facebook sites. An empirical analysis: The impact of different media and content types in Western Europe,” *Gov. Inf. Q.*, vol. 32, no. 1, pp. 52–62, Jan. 2015.
 - [61] Ø. Sæbø, J. Rose, and L. Skiftenes Flak, “The shape of eParticipation: Characterizing an emerging research area,” *Gov. Inf. Q.*, vol. 25, no. 3, pp. 400–428, Jul. 2008.
 - [62] A. Macintosh, T. F. Gordon, and A. Renton, “Providing Argument Support for E-Participation,” *J. Inf. Technol. Polit.*, vol. 6, no. 1, pp. 43–59, Feb. 2009.
 - [63] N. Karacapilidis, E. Loukis, and S. Dimopoulos, “Computer-supported G2G collaboration for public policy and decision-making,” *J. Enterp. Inf. Manag.*, vol. 18, no. 5, pp. 602–624, Oct. 2005.
 - [64] V. Peristeras, G. Mentzas, K. A. Tarabanis, and A. Abecker, “Transforming E-government and E-participation through IT,” *IEEE Intell. Syst.*, vol. 24, no. 5, pp. 14–19, Sep. 2009.
 - [65] S. A. Chun, S. Shulman, and R. Sandoval, “Government 2.0: Making Connections between Citizens, Data and Government.”
 - [66] M. A. Rezk, A. Ojo, and I. A. Hassan, “Mining Governmental Collaboration through Semantic Profiling of Open Data Catalogues and Publishers 1

- Introduction 2 Background and Related Work,” pp. 1–12.
- [67] L. Porwol, A. Ojo, and J. Breslin, “A semantic model for e-Participation.”
 - [68] L. Porwol, A. Ojo, and J. Breslin, “A semantic model for e-Participation - detailed conceptualization and ontology,” in *Proceedings of the 15th Annual International Conference on Digital Government Research - dg.o '14*, 2014, pp. 263–272.
 - [69] A. Ojo, T. Janowski, and E. Estevez, “Determining Progress Towards e-Government -What are the Core Indicators?,” 2007.
 - [70] L. Porwol, A. Ojo, and J. Breslin, “On the Duality of E-Participation – Towards a Foundation for Citizen-Led Participation,” Springer, Berlin, Heidelberg, 2013, pp. 211–225.
 - [71] L. Porwol, A. Ojo, and J. Breslin, “Harnessing the duality of e-participation,” in *Proceedings of the 7th International Conference on Theory and Practice of Electronic Governance - ICEGOV '13*, 2013, pp. 289–298.
 - [72] E. Kamateri *et al.*, “A Comparative Analysis of Tools and Technologies for Policy Making,” in *Policy Practice and Digital Science*, Cham: Springer International Publishing, 2015, pp. 125–156.
 - [73] M. Marrero, J. Urbano, S. Sánchez-Cuadrado, J. Morato, and J. M. Gómez-Berbís, “Named Entity Recognition: Fallacies, Challenges and Opportunities.”
 - [74] fao, “Better livestock policies for Africa,” 2016. [Online]. Available: <http://www.fao.org/wairdocs/ilri/x5499e/x5499e03.htm>. [Accessed: 29-Feb-2016].
 - [75] CDM – Executive Board, “Definitions of policy and programme of activities.” 2001.
 - [76] L. Siobhan, *Community Development Development*. 2006.
 - [77] C. L. Cochran and E. F. Malone, *Public Policy:Perspectives and Choices*. 2005.
 - [78] W3C, “VOCABULARIES,” 2015. [Online]. Available: <https://www.w3.org/standards/semanticweb/ontology>. [Accessed: 29-Mar-2017].
 - [79] T. R. Gruber, “A translation approach to portable ontology specifications,” *Knowl. Acquis.*, vol. 5, no. 2, pp. 199–220, 1993.
 - [80] D. Fensel, “Ontologies: A Silver Bullet for Knowledge Management and Electronic Systems,” *Berlin Heiderb. New York Springer-Varlag*, 2004.

- [81] R. W. Van Kralingen, P. R. S. Visser, T. J. M. Bench-Capon, and H. J. Van Den Herik, “A principled approach to developing legal knowledge systems,” *Int. J. Hum. Comput. Stud.*, vol. 51, no. 6, pp. 1127–1154, 1999.
- [82] A. Valente, “Types and roles of legal ontologies,” in *Law and the semantic web*, Springer, 2005, pp. 65–76.
- [83] H. Ben Sta, L. Ben Said, K. Ghédira, M. Bigand, and J. P. Bourey, “Cartographies of Ontology Concepts,” in *{ICEIS} 2005, Proceedings of the Seventh International Conference on Enterprise Information Systems, Miami, USA, May 25-28, 2005*, 2005, pp. 486–494.
- [84] M. E. Frisse *et al.*, “The financial impact of health information exchange on emergency department care,” *J. Am. Med. Informatics Assoc.*, vol. 19, no. 3, pp. 328–333, 2012.
- [85] J. D. Moffett and M. S. Sloman, “Policy hierarchies for distributed systems management,” *Sel. Areas Commun. IEEE J.*, vol. 11, no. 9, pp. 1404–1414, 1993.
- [86] J. M. Bradshaw *et al.*, “Making agents acceptable to people,” in *Intelligent Technologies for Information Analysis*, Springer, 2004, pp. 361–406.
- [87] L. Bunch *et al.*, “Human-agent teamwork in cyber operations: supporting co-evolution of tasks and artifacts with Luna,” in *Multiagent System Technologies*, Springer, 2012, pp. 53–67.
- [88] R. Montanari, E. Lupu, and C. Stefanelli, “Policy-based dynamic reconfiguration of mobile-code applications,” *Computer (Long. Beach. Calif.)*, vol. 37, no. 7, pp. 73–80, 2004.
- [89] J. Keeney and V. Cahill, “Chisel: A policy-driven, context-aware, dynamic adaptation framework,” in *Policies for Distributed Systems and Networks, 2003. Proceedings. POLICY 2003. IEEE 4th International Workshop on*, 2003, pp. 3–14.
- [90] E. Lupu, N. Dulay, J. Sventek, and M. Sloman, “Autonomous pervasive systems and the policy challenges of a small world!,” in *Policies for Distributed Systems and Networks, 2007. POLICY'07. Eighth IEEE International Workshop on*, 2007, pp. 3–7.
- [91] G. Xu, C. Borcea, and L. Iftode, “A policy enforcing mechanism for trusted ad hoc networks,” *Dependable Secur. Comput. IEEE Trans.*, vol. 8, no. 3, pp. 321–

- 336, 2011.
- [92] N. Henry, *Public administration and public affairs*. Routledge, 2015.
 - [93] E. Ostrom, *Understanding institutional diversity*. Princeton university press, 2009.
 - [94] P. Hassenteufel, *Sociologie politique: l'action publique*. A. Colin Paris, 2008.
 - [95] P. André, C. E. Delisle, and J.-P. Revéret, *Environmental assessment for sustainable development: Processes, actors and practice*. Presses inter Polytechnique, 2004.
 - [96] P. Mazzega and R. Boulet, “Méthodologie d’Analyse Réticulaire d’une Politique Publique: le SDAGE Adour-Garonne 2010-2015,” *Polit. Publiques Systèmes Complexes*, pp. 978–27056, 2012.
 - [97] European Commission, “Impact Assessment: Next Steps In support of competitiveness and sustainable development,” 2004.
 - [98] D. OEC, “Organization for Economic Cooperation & Development,” *Econ. Surv. United States*, vol. 2008, 2008.
 - [99] DEMO-net Consortium, “D 5.2 eParticipation: The potential of new and emerging technologies.,” 2007. [Online]. Available: <http://www.demo-net.org>.
 - [100] E. Young and L. Quinn, “Writing effective public policy papers,” 2002.
 - [101] A. Ranney, “Political science and public policy,” 1968.
 - [102] K. Kim, “Public Policy and Governance: Some Thoughts on Its Elements,” Available SSRN 2589526, 2015.
 - [103] C. M. Martin *et al.*, “Avoidable Hospitalizations in Older Adults,” in *Handbook of Systems and Complexity in Health*, Springer, 2013, pp. 445–465.
 - [104] M. Mackay and L. Shaxton, “Understanding and Applying Basic Public Policy Concepts.” University of Guelph, 2005.
 - [105] S. J. Ball, M. Maguire, and A. Braun, *How schools do policy: Policy enactments in secondary schools*. Routledge, 2012.
 - [106] A. Braun, S. J. Ball, and M. Maguire, “Policy enactments in schools introduction: Towards a toolbox for theory and research,” *Discourse Stud. Cult. Polit. Educ.*, vol. 32, no. 4, pp. 581–583, 2011.
 - [107] B. B. Bernstein, *Pedagogy, symbolic control, and identity: Theory, research, critique*, no. 4. Rowman & Littlefield, 2000.
 - [108] S. J. Ball, *Education policy and social class: The selected works of Stephen J.*

Ball. Routledge, 2005.

- [109] M. A. Wimmer, K. Furdik, M. Bicking, M. Mach, T. Sabol, and P. Butka, “Open Collaboration in Policy Development: Concept and Architecture to integrate scenario development and formal policy modelling,” in *Empowering open and collaborative governance*, Springer, 2012, pp. 199–219.
- [110] T. N. Tsekos, “Towards Integrated Policy Making: Remedyng the Public Action Dichotomy Through Information and Communication Technologies and Learning,” in *State Modernization and Decentralization: Implications for Education and Training. Proceedings of the joint IIASIA-NISPacee Conference Public Administration Education and Training in a Framework of Modernization and Decentralization: United Natons, NISPACe*, 2003.
- [111] C. Gray, “The Structure of Cultural Policy.” University of Warwick, UK.
- [112] C. West and C. H. F. Smith, “WE ARE NOT A GOVERNMENT POODLE 1: Museums and social inclusion under New Labour,” *Int. J. Cult. policy*, vol. 11, no. 3, pp. 275–288, 2005.
- [113] C. Gray, “Analysing cultural policy: incorrigibly plural or ontologically incompatible?,” *Int. J. Cult. policy*, vol. 16, no. 2, pp. 215–230, 2010.
- [114] M. S. Archer, *Culture and agency: The place of culture in social theory*. Cambridge University Press, 1996.
- [115] M. S. Archer, *Realist social theory: The morphogenetic approach*. Cambridge university press, 1995.
- [116] M. S. Archer, *Structure, agency and the internal conversation*. Cambridge University Press, 2003.
- [117] D. ELDER-VASS, “For emergence: refining Archer’s account of social structure,” *J. Theory Soc. Behav.*, vol. 37, no. 1, pp. 25–44, 2007.
- [118] A. Olivé and J. Cabot, “A research agenda for conceptual schema-centric development,” in *Conceptual Modelling in Information Systems Engineering*, Springer, 2007, pp. 319–334.
- [119] C. Rolland, “Capturing system intentionality with maps,” in *Conceptual modelling in Information Systems engineering*, Springer, 2007, pp. 141–158.
- [120] P. Johannesson, “The role of business models in enterprise modelling,” in *Conceptual modelling in information systems engineering*, Springer, 2007, pp. 123–140.

- [121] Y. Wand and R. Weber, “Research commentary: information systems and conceptual modeling a research agenda,” *Inf. Syst. Res.*, vol. 13, no. 4, pp. 363–376, 2002.
- [122] Y. Wand, D. E. Monarchi, J. Parsons, and C. C. Woo, “Theoretical foundations for conceptual modelling in information systems development,” *Decis. Support Syst.*, vol. 15, no. 4, pp. 285–304, 1995.
- [123] C. Rolland, “Information Systems and Web Information Systems: A Methodological Perspective,” in *Information Technology and Communication at the Dawn of the New Millennium*, 2000, p. 1.
- [124] J. A. Bubenko Jr, “From information algebra to enterprise modelling and ontologies A historical perspective on modelling for information systems,” in *Conceptual Modelling in Information Systems Engineering*, Springer, 2007, pp. 1–18.
- [125] T. W. Olle, H. G. Sol, and I. G. MacDonald, *Information Systems Methodologies; A Framework for Understanding*. Addison-Wesley Longman Publishing Co., Inc., 1991.
- [126] J. Mylopoulos, L. Chung, and E. Yu, “From object-oriented to goal-oriented requirements analysis,” *Commun. ACM*, vol. 42, no. 1, pp. 31–37, 1999.
- [127] C. Rolland and N. Prakash, “From conceptual modelling to requirements engineering,” *Ann. Softw. Eng.*, vol. 10, no. 1–4, pp. 151–176, 2000.
- [128] I. Davies, P. Green, M. Rosemann, M. Indulska, and S. Gallo, “How do practitioners use conceptual modeling in practice?,” *Data Knowl. Eng.*, vol. 58, no. 3, pp. 358–380, 2006.
- [129] J. Webster and R. T. Watson, “Analyzing the past to prepare for the future: Writing a,” *MIS Q.*, vol. 26, no. 2, pp. 13–23, 2002.
- [130] E. Kaliva, E. Panopoulou, E. Tambouris, and K. Tarabanis, “Comparative Analysis of Conceptual and Domain Models,” *White Pap. Comp. Anal. performed Performed by eGovPoliNet partners (Annexes to D4.2)*, no. Figure 1, pp. 1–14, 2014.
- [131] G. Ciasullo, G. Lodi, and A. Rotundo, “Core Public Service Vocabulary : The Italian Application Profile,” *Agency Digit. Italy*, 2014.
- [132] B. Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, “Systematic literature reviews in software engineering--a systematic

- literature review,” *Inf. Softw. Technol.*, vol. 51, no. 1, pp. 7–15, 2009.
- [133] P. A. Wc, “Piloting the Core Public Service Vocabulary ‘ Describe your public service once to publish on multiple Government Access Portals ’ Document Metadata,” p. 45, 2013.
- [134] P. Archer, S. Goedertier, and N. Loutas, “Study on persistent URIs, eith identification of best practices and recommendations on the topic for the MSs and the EC,” p. 45, 2012.
- [135] M. Janssen, Y. Charalabidis, and A. Zuiderwijk, “Benefits, Adoption Barriers and Myths of Open Data and Open Government,” *Inf. Syst. Manag.*, vol. 29, no. 4, pp. 258–268, 2012.
- [136] M. Kassen, “A promising phenomenon of open data : A case study of the Chicago open data project,” *Gov. Inf. Q.*, vol. 30, no. 4, pp. 508–513, 2013.
- [137] R. Grishman, “Message Understanding Conference-6: A Brief History.,” *Proc. COLING*, vol. 96, 1996.
- [138] A. McCallum and W. Li, “Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons,” in *Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003 -*, 2003, vol. 4, pp. 188–191.
- [139] D. M. Bikel, S. Miller, R. Schwartz, and R. Weischedel, “Nymble,” in *Proceedings of the fifth conference on Applied natural language processing -*, 1997, pp. 194–201.
- [140] A. Borthwick and J. Sterling, “NYU: Description of the MENE named entity system as used in MUC-7,” ... *Conf. (MUC-7*, 1998.
- [141] M. Asahara and Y. Matsumoto, “Japanese Named Entity extraction with redundant morphological analysis,” in *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - NAACL '03*, 2003, vol. 1, pp. 8–15.
- [142] J. Hoffart *et al.*, “Robust disambiguation of named entities in text,” pp. 782–792, Jul. 2011.
- [143] H. Ji and R. Grishman, “Data selection in semi-supervised learning for name tagging,” pp. 48–55, Jul. 2006.
- [144] E. Alfonseca and S. Manandhar, “An unsupervised method for general named entity recognition and automated concept discovery,” ... *Conf. Gen.* ..., 2002.

- [145] C. H. Ku, A. Iribarri, G. Leroy, and D. Ph, “Natural Language Processing and e-Government: Crime Information Extraction from Heterogeneous Data Sources,” in *The proceedings of the 9th Annual International Digital Government Research Conference*, 2006, pp. 162–170.
- [146] H. Dalianis, M. Rosell, and E. Sneiders, “Clustering E-Mails for the Swedish Social Insurance Agency – What Part of the E-Mail Thread Gives the Best Quality ?,” pp. 115–120, 2010.
- [147] F. Amato, a. Mazzeo, V. Moscato, and a. Picariello, “Semantic Management of Multimedia Documents for E-Government Activity,” *2009 Int. Conf. Complex, Intell. Softw. Intensive Syst.*, pp. 1193–1198, Mar. 2009.
- [148] L. M. Williams, S. A. Cody, and J. Parnell, “Prospecting for new collaborations: Mining syllabi for library service opportunities,” *J. Acad. Librariansh.*, vol. 30, no. 4, pp. 270–275, 2004.
- [149] R. Basanya, A. Ojo, T. Janowski, and F. Turini, “Mining collaboration opportunities to support Joined-Up Government,” *IFIP Adv. Inf. Commun. Technol.*, vol. 362 AICT, pp. 359–366, 2011.
- [150] L. Wan, J. Chen, and D. Gu, “An Information Mining Model of Intelligent Collaboration Based on Agent Technology,” in *International Conference on Applied Sciences, Engineering and Technology, ICASET 2014*, 2014.
- [151] C. Palmer, J. A. Harding, R. Swarnkar, B. P. Das, and R. I. M. Young, “Generating rules from data mining for collaboration moderator services,” *J. Intell. Manuf.*, vol. 24, no. 2, pp. 313–330, 2013.
- [152] The Open Knowledge Foundation, “Ckan - The open source data portal software.” .
- [153] A. K. McCallum, “MALLET: The Machine Learning for Language Toolkit.” .
- [154] S. E. Whang, D. Marmaros, and H. Garcia-Molina, “Pay-as-you-go entity resolution,” *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 5, pp. 1111–1124, 2013.
- [155] D. Lin, “Automatic retrieval and clustering of similar words,” in *Proceedings of the 17th international conference on Computational linguistics-Volume 2*, 1998, pp. 768–774.
- [156] H. Matthew, “University of Manchester OWL Validator,” 2009. .
- [157] W3C, “RDF Validator,” 2001. .

- [158] W3C, “SWValidators,” 2011..
- [159] C. Bezerra, F. Freitas, and F. Santana, “Evaluating ontologies with Competency Questions,” *Proc. - 2013 IEEE/WIC/ACM Int. Jt. Conf. Web Intell. Intell. Agent Technol. - Work. WI-IATW 2013*, vol. 3, pp. 284–285, 2013.
- [160] M. Gruninger and M. S. Fox, “Methodology for the Design and Evaluation of Ontologies,” *Ind. Eng.*, vol. 95, pp. 1–10, 1995.
- [161] M. A. Rezk, A. Ojo, G. A. El Khayat, and S. Hussein, “A Government Decision Analytics Framework Based on Citizen Opinion,” in *9th International Conference on Theory and Practice of Electronic Governance*, 2016.
- [162] U. Government, “UK Bioenergy Strategy,” 2012.
- [163] H. E. Authority, “National Plan for Equity of Access to Higher Education 2015-2019,” 2015.
- [164] F. Guillet and H. J. Hamilton, *Quality measures in data mining*, vol. 43. Springer, 2007.

APPENDIX A

CPPV:

Documentation: <http://vmogi01.deri.ie/egoveppv/>

Live instance: <http://vmrtpa05.deri.ie:8080/dataset>

Code: <https://github.com/mohadelrezk/cppv-ckan-implementation>

APPENDIX B

Policy Collaboration Mining:

Code: <https://github.com/mohadelrezk/opengovcollaborationanalysis>

APPENDIX C

GOVDAF:

Code: <https://github.com/mohadelrezk/GOV-DAF>

Code: <https://github.com/mohadelrezk/policyanalytics>