MIC 2015
Management International Conference
Portorož, Slovenia • 28–30 May 2015

Sorting Through Waste Management Literature: A Text Mining Approach to a Literature Review

Ksenia Silchenko
University of Macerata, Department of Economics and Law, Italy
k.silchenko@unimc.it

Roberto Del Gobbo
University of Macerata, Department of Economics and Law, Italy
roberto.delgobbo@unimc.it

Nicola Castellano
University of Macerata, Department of Economics and Law, Italy
nicola.castellano@unimc.it

Bruno Maria Franceschetti
University of Macerata, Department of Economics and Law, Italy
bruno.franceschetti@unimc.it

Virginia Tosi
University of Macerata, Department of Economics and Law, Italy
v.tosi@unimc.it

Monia La Verghetta
University of Macerata, Department of Economics and Law, Italy
monia.laverghetta@unimc.it

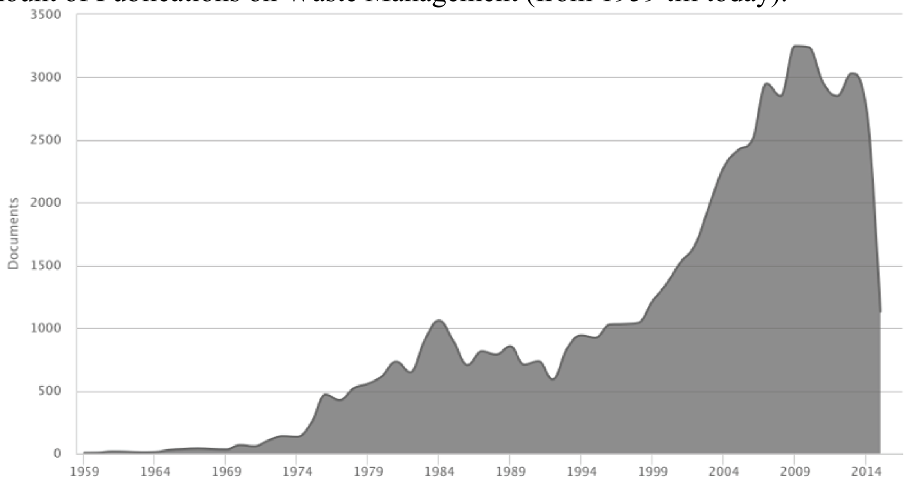
Abstract. With sustainability and management of waste as focus of multiple disciplines, there is still a considerable gap in the academic literature in regard to the definition of "waste management". The present research addresses the issue of a substantial lack of an acceptable interdisciplinary definition of waste management by means of synthesizing existing literature on the matter and identifying the most recurrent and relevant waste management concepts by applying the method of text mining. The results allow gaining a deeper understanding of (1) the typical concepts of each scientific discipline that studies waste management, (2) cross-disciplinary concept differences and similarities, and (3) the concept networks that can become potential building blocks of the waste management studies definition. Finally, a number of future research directions and propositions are suggested.

Keywords: waste management, literature review, text mining, network analysis.

1 Introduction

*"*Garbage is a great resource in the wrong place lacking someone's imagination to recycle it into everyone's benefit*"* (Hansen 2015). While still in the wrong place, "waste" certainly generates the right level of attention among scientists, policymakers, business prefessionals, and regular citizens all over the globe. Accordingly, the amount of publications on "waste management" topics has grown exponentially. For instance, the keyword search of "waste management" on Scopus online bibliographic database results in 58.746 publications between academic peer-reviewed articles, trade news and institutional documents published from 1959 until today. The growth of the interest is shown on figure 1 below.

Figure 1: Amount of Publications on Waste Management (from 1959 till today).



Adapted from Scopus electronic bibliographic database, "waste management" search query, retrieved on July 20, 2015.

In 1959 the interest to waste management research is almost inexistent counting only 5 publications, but in 2010 it reaches the peak with 3.239 publications. Even though academic researchers, policy makers and professionals show more and more concern with waste and its management, up till today there is no comprehensive and shared definition of what waste management is. Moreover, the complexity and practical importance of managing waste makes it crucial to employ an interdisciplinary perspective to its research.

Based on the above, the current research is aimed at studying the existing waste management literature in order to shed light on various conceptual connections within waste management research.

The literature review was conducted following the method described by Cooper (1988) and the findings were further scrutinized via text mining technique in order to identify the most recurrent concepts within studies of waste management.

The results help better understand what are:
(1) the typical concepts of each scientific discipline that studies waste management. A taxonomy based on the level of concepts' interdisciplinarity revealed a number of general, overarching, common and specific waste management concepts;
(2) the cross-disciplinary differences and similarities of key recurrent concepts;
(3) the concept networks that can become potential building blocks of the waste management studies definition.

The conducted literature review confirms the absences of a clear and shared definition of waste management. Even though finding or creating the definition is not the direct objective of the present research, it might become a goal for further investigation.

This paper is structured as following: section 2 describes the literature review selection criteria and protocol, section 3 focuses on the text mining technique employed to analyse the data collected during the literature review, and section 4 discusses the findings. Finally, the conclusions emphasize summarizing remarks, limitations and opportunities for future research.

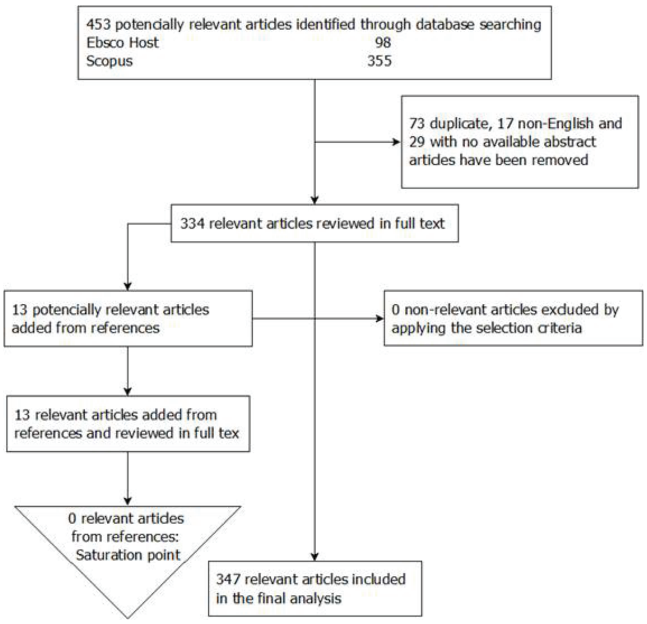
2 Waste management literature: Selecting relevant publications

A systematic literature review on waste management was conducted following methodology outlined by Cooper (1988), choosing an option of a selective approach to a keyword search of peer reviewed scholarly articles. Two major electronic bibliographic databases were consulted to locate and select the publications. Then, a group of 14 researchers collectively worked with the sample in order to create, pilot-test and confirm the coding protocol, which was later used to analyse all the publications in a comparable and uniform manner. Finally, the selected publications were classified by their scientific affinity following European Research Council (ERC) taxonomy. Such categorization served to proceed to a further step of the analysis via text mining.

2.1 Database search and selection strategy

In order to retrieve relevant publications in the field of waste management research, two major electronic bibliographic databases, Ebsco Host-Business Source Complete and Scopus, were consulted. The search query included keyword "waste management" linked via Boolean AND with each of the following keywords: "state of the art", “literature”, and “defin*”. The search conducted in January 2014 resulted in overall 453 articles. Figure 2 shows the screening process employeed to the results retrieval.

Figure 2. Flow Diagram for Literature Selection Process



First, duplicates (73 articles) and studies with no available abstract (29) were discarded. For practical reasons, only studies in English were taken into consideration and thus non-English (17) articles were eliminated from the final pool.

According to Randolph (2009), electronic searches may lead to an insufficient amount of articles for a thematically-exhaustive review and as suggested by the author “*the most effective method may be to search the references of the articles that were retrieved, determine which of those seem relevant, find those, read their references, and repeat the process until a point of saturation is reached — a point where no new relevant articles come to light*” (Randolph 2009, 7). References retrieved from the citations were therefore used as a secondary, but essential, source. After iterative cross-referencing of 334 articles, 13 additional articles were found and included to the sample. This search strategy identified a total of 347 qualifying articles, published between 1976 and 2014.

2.2 Coding frame: a formal reading protocol

In order to read, classify and analyse the content of the retrieved publications, a formal protocol was developed with the help of a team of 14 researchers. Table 1 schematically represents the final version, or a coding frame, of the formal reading protocol.
The coding frame was developed gradually and tested by the entire group of researchers involved in the period of May-July 2014. First drafts of the coding frame were tested on a sample of 1-3 articles per researcher. Regular meetings allowed on-going discussion about potential difficulties and ambiguities of the information to be captured and criteria for coding, which eventually led to several modifications of the protocol. In the end of group negotiation and testing, all the researchers used the same coding frame that they followed in the analysis of the assigned articles. This coding frame was supplied with detailed instructions on the format of coding (e.g. free text or "yes/no" choice), the amount of detail to capture, and the guidelines how to handle ambiguities. All the individual difficulties were discussed and addressed on a case-by-case basis.

Table 1. Literature Review Protocol: Coding Frame Categories.

Bibliographic data	**ERC domain**	**Research origin**	**Article content**	**Research methodology**	**Waste management**
Author(s); Title; Journal; Year/Nr; Keywords; Abstract; Number of Google scholar citations	Macro domain (e.g. SH); Discipline (e.g. SH1); Sub-discipline (e.g. SH1_3); Comments	Authors' country; Authors' institution type (e.g. University, Public institution, Private company); Authors' institution; Research data country	Research objective; Research results; Audience (specialized scholars, general scholars, practitioners or policy makers, general public)	Applied method; Method type (Qualitative, Empirical research, Quantitative descriptive, Quantitative inferential)	Type of waste; Definition of "waste management"; Related definitions

The reading and analysis of the articles took place in June-October 2014. In a few cases, where it was impossible to retrieve full texts of the papers, the analysis was based on reading the abstract.

2.3 Thematic grouping by scientific domains (ERC)

The results were aggregated, cleaned and homogenised by the lead team composed of the authors of the present study. Table 2 shows the detailed synthesis of the analysed articles per their scientific affinity (ERC domain), which was considered the most significant criteria to categorize the articles in our sample.

Table 2: Literature review synthesis, by articles' scientific affiliation (ERC domain).

		SH: Social Sciences and Humanities 73%, n=256							
		SH1 Individuals, institutions and markets 20%, n=70	**SH2 Institutions, values, beliefs and behaviour** 15%, n=53	**SH3 Environment and society** 37%, n=129		**PE: Mathematics, physical sciences, information and communication, engineering, universe and earth sciences** 21%, n=71	**LS: Life Sciences** 4%, n=13	**Other** 3%, n=11	**Total** 100%, n=347
Authors' origin:	Europe	55%	93%	55%	60%	58%	55%	60%	65%
	North America	29%	13%	21%	23%	15%	-	50%	22%
	Asia Pacific	28%	3%	27%	23%	25%	36%	10%	26%
	Africa	9%	-	7%	6%	5%	9%	-	6%
	South America	9%	-	2%	2%	5%	9%	-	3%
Author's affiliation:	University	91%	79%	85%	86%	95%	73%	90%	80%
	Public Institution	9%	10%	16%	13%	8%	18%	10%	7.4%
	Private company	6%	17%	9%	9%	5%	9%	-	4%
Research approach:	Qualitative	74%	93%	67%	74%	44%	18%	56%	55%
	Quantitative	47%	13%	49%	42%	41%	91%	33%	31%
	Empirical data analysis	53%	37%	53%	50%	58%	82%	56%	53%
Research data origin:	Europe	41%	82%	51%	52%	4%	45%	67%	47%
	North America	15%	23%	21%	20%	17%		50%	17%
	Asia Pacific	37%	5%	26%	26%	39%	36%	17%	26%
	Africa	9%	-	11%	9%	7%	9%	17%	8%
	South America	4%	-	3%	3%	7%		17%	3%
Average # of Google citations:		18,95	10,28	19,9	18,07	18,07	11,55	25,07	18,55
Top types of waste:		Urban and solid waste, general/ multiple type waste, construction and demolition waste, hazardous & radioactive, e-waste	Waste in general, municipal solid waste, EU waste, hazardous waste, radioactive waste	General/multiple types waste, solid urban and solid waste, industrial waste, e-waste, hazardous waste	General and multiple types waste, solid waste, solid urban waste, hazardous waste, e-waste, construction and demolition waste, industrial waste	General and multiple types waste, solid waste, solid urban waste, hazardous waste, e-waste, construction and demolition waste, industrial waste	Municipal waste, organic waste, solid waste, pathological waste	Solid waste, metal waste, concept of waste, waste as a process, constructio n waste	General/ multiple waste, solid and urban waste, construction & demolition waste, hazardous waste; e-waste, industrial waste

The synthesis of the results confirmed that the number of publications has increased exponentially during the past 15 years. Only about 13.5% of the studies were published before the year of 2000. The rest of the publications in the sample were almost equally distributed between the first decade of 2000s (44.5%) and only five years of 2010s (42%).

Scholarly articles written by researchers with University credentials (80%) prevail in the sample and, in some cases, the articles are authored by public policymakers (7.4%) or professionals (4%) from the private sector. At the same time, the intended audience of the studies is not necessarily academic. Even though scholar audience (84%) could benefit from the majority of studies, a good number of studies (73%) are destined for waste management professionals and public officials, and some - forgeneral public (20%).

The most frequently used keywords were found to be: "waste management", "life cycle assessment", "sustainability", "environment", "solid waste", "reverse logistics", "industrial ecology" and "recycling".

While one journal in our sample ("Resources, Conservation and Recycling") could be considered the leading publication outlet of the literature on waste management accounting for about 10% of the selected articles, there is a very long tail of journals that had no more than 8-9 (and most frequently only 1-3) articles published.

In our sample the top countries of authors' universities or institutions are: USA (45), UK (44), Italy (24), China (14), Canada (13) and India (10). Even though selection of the language (English) could bias our results by putting four English-speaking countries in the top list, overall representation of European authors combined altogether outpaces scholars from other continents and the overall country of origin mix is quite heterogeneous. Interestingly, while UK and US data are those more often used in the studies (naturally, related to the country of researchers' origin as shown before), EU data (i.e. collective of several EU countries) are analysed extensively, as our results show.

Methodology-wise, approximately 55% of the articles analysed are designed as qualitative studies, 31% - as quantitative, 14% - as mixed, and approximately 53% of them relied on the use of empiric data.

Overall we found a high level of heterogeneity in almost every analysed field. In order to reach a higher level of clarity, it was decided to treat the entire sample by taking into consideration which scientific discipline a particular study belongs to. As it's shown in Table 2, 21% of the articles belong to PE (Physical Sciences and Engineering) macro-field, 73% - to SH (Social Sciences and Humanities) macro-field, which can be further broken down to SH1 (Economics, Finance and Management) - 20%, SH2 (Sociology, social studies, political science, law and communication) - 15%, and SH3 (Environmental studies, demography, social geography, urban and regional studies) - 37%. Other smaller groups included other SH (SH4, SH5) - 1%, LS (Life Sciences) - 4%, and some interdisciplinary studies (PE and SH) 2%.

Classification of the selected publications according to ERC scientific domains is fundamental for the objectives of the present research, which aims to identify disciplinary differences and similarities in key concepts employed in the studies of waste management. However, in order to guarantee significance of the results in the following steps of the analysis via text mining we had to ensure that each segment of articles (grouped by ERC domain principle) had a sufficient number of texts. As a result, only 4 segments (SH1, SH2, SH3, PE) were promoted to the further steps of the analysis while the remaining segments, accounting for about 7% of the sample overall, were discarded.

3 Data analysis: text mining approach

Text mining approach, just like the broader family of methods of data mining, could be defined as a process of generating knowledge through elaboration of large samples of documents and databases (Tan 1999). In case of text mining, we are talking exclusively about analysis of textual data. Some scholars consider text mining as a strategically powerful technique allowing extraction of relevant insights from large unstructured sets of data, thus turning "hidden data" into ordered sets of semantic and conceptual information (Bolasco and Canzonetti 2003; Dulli, Polpettini and Trotta 2004).

Text mining approach to a systematic waste management literature review was applied to the texts of 347 abstracts following three steps. First, "distinctive words" were identified for each of four segments grouped by their scientific discipline. Second, distinctive words were transformed into relevant concepts by taking into consideration most frequent word combinations used in the texts. Third, the relationships between various concepts were analysed for each scientific discipline and for the entire sample. The analysis was conducted with the help of KHCoder software.

3.1 Domain-specific vocabulary: Distinctive words

The first step of text mining focused on identifying specific vocabulary for each ERC domain. To do so the frequency of words use was counted based on the analysis of text of the publication's title, abstract and keywords. Only nouns and adjectives were taken into consideration in order to preserve linguistic and conceptual significance, thus eliminating verbs. Lexical or textual analysis as a rule relies on lemmas as a unit of analysis. Lemma is a canonical or dictionary form of a word chosen by convention to represent all word forms (Leopold and Kindermann 2002). In case of verbs, lemma is usually infinitive (Guerin-Pace 1998), which makes it difficult to operate via text mining on the level of word combinations and concepts. To standardize and simplify the basic units of the analysis, the verbs were excluded.

The measure of conditioned probability helped to identify whether or not (and how much) frequently used words were specific to the analysed ERC domains (Miner et al 2012). As explained before, each of four segments of publications grouped on the basis of ERC domain/discipline was analysed separately in order to identify the most recurrent words first. To add more rigor, the analysis took into consideration not only the "absolute" frequency, but specificity of vocabulary for each scientific discipline or "relative frequency" (Egghe and Michel 2002). The "distinctive" words were identified using the similarity index, namely Jaccard index (Huang 2008), calculated as ratio between $A \cap B$ "intersection" probability and $A \cup B$ "union" probability, where A is a certain word and B is a segment of ERC domain/discipline.

The index represents the ratio between: i) the probability that the word is used in the texts of one scientific domain, ii) the sum of probability that the word can be used in all the texts, and iii) the proportion of the texts of a specific domain in relation to all the texts. Mathematically, it can be expressed as formula (1) below:

(1) $$Jaccard\ index = \frac{A}{A+B+C}$$

where A stands for the number of documents belonging to a specific scientific domain/discipline where the word is used; B – for the total number of documents where the word is used; and C – for the number of documents belonging to a specific scientific domain/discipline.

Table 3 shows the list of the distinctive words for each of four ERC domains/disciplines. Some words were eventually excluded from the final selection (bottom part of table 3) due to the fact that they were very general (specific to all management and/or academic literature, e.g. "study", "literature", "result") and could not contribute to the specific analysis of waste management studies.

Table 3: Top 10 distinctive words per ERC domain/discipline.

SH1			SH2			SH3			PE		
Distinctive word	*POS**	*Jaccard*	*Distinctive word*	*POS**	*Jaccard*	*Distinctive word*	*POS**	*Jaccard*	*Distinctive word*	*POS**	*Jaccard*
Selected distinctive words											
management	*Noun*	0.2172	definition	*Noun*	0.2222	environmental	*Adj*	0.2167	process	*Noun*	0.1373
waste	*Noun*	0.1922	waste	*Noun*	0.1657	policy	*Noun*	0.1423	material	*Noun*	0.1320
policy	*Noun*	0.1656	law	*Noun*	0.1585	material	*Noun*	0.1412	cost	*Noun*	0.1314
environmental	*Adj*	0.1626	environmental	*Adj*	0.1188	product	*Noun*	0.1339	treatment	*Noun*	0.1266
framework	*Noun*	0.1500	control	*Noun*	0.1170				model	*Noun*	0.1257
			new	*Adj*	0.1167						
Other											
literature	*Noun*	0.2312	article	*Noun*	0.1927	study	*Noun*	0.2077	result	*Noun*	0.1584
study	*Noun*	0.1568	EU	*Noun*	0.1429	use	*Noun*	0.1393	study	*Noun*	0.1509
result	*Noun*	0.1529	provision	*Noun*	0.1250	datum	*Noun*	0.1355	application	*Noun*	0.1220
different	*Adj*	0.1479	relation	*Noun*	0.1190	approach	*Noun*	0.1328	impact	*Noun*	0.1217
general	*Adj*	0.1429				problem	*Noun*	0.1323	different	*Adj*	0.1192
						collection	*Noun*	0.1288			

**POS – part of speech*

3.2 From words to concepts: analysis of concordance

The objective of the following step is to build "recurrent concepts" from the combination of selected "distinctive words" identified in the previous step and most frequent word combinations with them. As before, the analysis was conducted for each segment grouped by ERC domain/discipline. As explained by Bolasco (2005) such analysis defined as "analysis of concordance", as an output produces a body of "co-texts" with node words (derived from "distinctive words") and most frequently used words in the immediate left or right positions within texts (max. 5 positions before or after the node word).

As a synthetic concordance measure used to identify the most significant word combinations could be expressed via a score presented in a formula (2) below:

(2) $$S(w) = \sum_{i=1}^{5} \frac{(l_i + r_i)}{i}$$

where l_i stands for the frequency of a word w occurrence i-number of words before (i.e. on the left) from the node word. On the other hand, r_i stands for the frequency of a word w occurrence i-number of words after (i.e. on the right) from the node word. The higher frequency of a certain word w concordance with the node word on the left or on the right $(l_i + r_i)$ - the higher *S(w)* score it will return. Calculating the *S(w)* score involves taking into consideration the fact that concordance $(l_i + r_i)$ depends on the distance i between node words and precedent/following words: shorter distance produces a higher score.

The list of word combinations for each node word was ordered based on the *S(w)* score and, after a linguistic check, first 10 "valid" results were chosen. Table 4 shows the final list of concepts derived from the most frequent word combinations of node words and words on their immediate left or right.

Table 4: Concepts Derived from Node Word Combinations, per ERC domain/discipline.

SH1 (economics, finance and management)				
environmental	**framework**	**management**	**policy**	**waste**
environmental performance	regulatory framework	waste management	environmental policy	waste management
environmental policy	theoretical framework	management practice	waste policy	construction waste
environmental management	management framework	environmental management	policy frame	solid waste
environmental impact	research framework	performance management	public policy	demolition waste
environmental protection	contextual framework	management education	management policy	waste service
environmental issue	institutional framework	management policy	economic policy	waste minimization
environmental regulation	legislative framework	management theory	policy deliberation	waste indicator
environmental practice	modelling framework	e-waste management	policy idea	waste collection
environmental technology	quantitative framework	risk management	policy maker	waste reduction
environmental cost	stochastic framework	water management	policy tool	waste policy

SH2 (sociology, social anthropology, political science, law, communication, social studies of science and technology)					
environmental	**control**	**definition**	**law**	**waste**	**new**
environmental assessment	waste control	legal definition	waste law	waste management	new law
environmental protection	mandatory control	broad definition	EC law	waste disposal	new waste
environmental impact	applicable control	clear definition	new law	waste policy	new problem
environmental policy	democratic control	EC definition	Community law	waste treatment	new act
environmental concern	disease control	alternative definition	Merli law	waste law	new element
environmental conflict	legislative control	central definition	anti-trust law	radioactive waste	new guideline
environmental protection	control officer	complete definition	applicable law	waste directive	new insight
environmental benefit	pollution control	directive definition	law certainly	Community waste	new investment
environmental pollution	control procedure	overarching definition	Delaware law	EC waste	
environmental management	voluntary control	precise definition	International law	waste regulation	

SH3 (environmental studies, demography, social geography, urban and regional studies)			
environmental	**material**	**policy**	**product**
environmental activity	material flow	environmental policy	environmental product
environmental assessment	waste material	policy goal	waste product
environmental impact	material recovery	policy network	product life
environmental issue	raw material	waste policy	product policy
environmental management	recyclable material	disposal policy	green product
environmental performance	secondary material	policy implication	integrated product
environmental policy	alternative material	management policy	intermediate product
environmental product	combustible material	product policy	PVC product
environmental protection	material cycle	policy maker	product stewardship
environmental strategy	construction material	development policy	product management
product environmental			

PE (Mathematics, physical sciences, information and communication, engineering, universe and earth sciences)				
cost	**material**	**model**	**process**	**treatment**
energy cost	programming model	BWAS model	unit process	waste treatment
disposal cost	I-O-W model	cost model	production process	treatment option
management cost	management model	input-output model	reuse process	treatment plant
care cost	cost model	I-O-W model	assessment process	alternative treatment
investment cost	input-output model	management model	LCA process	end-of-life treatment
cost model	mathematical model	mathematical model	composting process	water treatment
cost reduction	thinking model	process model	decision process	treatment technology
transportation cost	quality model	programming model	chemical process	end-of-pipe treatment
material cost	process model	quality model	industrial process	treatment facility
operating cost	BWAS model	thinking model	process impact	treatment infrastructure
significant cost				

3.3 Connections between concepts: Betweenness centrality and co-occurrence networks

The objective of the third and final stage of text mining consists in identifying existing connections between various concepts or, in other words, creation and visualization of co-occurrence network. Figures 3 and 4, discussed in more detail in the following sections, show the co-occurrence networks, with and without taking into consideration ERC domain/disciplines respectively.

The method developed by Fruchterman & Reingold (1991) helped to arrange the concepts in a form of a network or a map, which visually aids reading and understanding the results. Nodes represent the concepts, while lines – co-occurrences of concepts within texts (Özgür et al. 2008). Stronger and more frequent co-occurrences between concepts, calculated via Jaccard index, are depicted with bolder lines. The length of lines and the proximity of nodes, on the other hand, are purely arbitrary and do not necessarily represent a closer conceptual association or stronger co-occurrence.

The overall co-occurrence network on figure 4 shows only the strongest co-occurrences without taking into consideration ERC domains/disciplines. The strength of association is measured for each possible combination between the concepts included in the analysis. To define the centrality of the nodes, a measure called "betweenness centrality" was applied. It's based on the frequency of each single node occurrence within the shortest path connecting various concepts (Freeman 1977).

Betweenness centrality indicates how each node is connected to other nodes in-between the network. In other words, betweenness centrality measures the relevance of the node by evaluating its presence in the connection paths between various nodes. Formally speaking, betweenness of node *X* equals to number of paths of the minimal length for all the origin-destination node combinations that include node *X*, normalized according to the maximum number of possible combinations. The mathematical count therefore involves the following:

1. identify all node combination (couple of origin-destination);
2. identify the minimal length connection path(s) between the nodes for each couple;
3. count the number of connection paths that involve a specific node and exclude all the connection paths where origin and destination are the same node;
4. calculate the total number of connected node couples (discarding those excluded in step 3)
5. normalize the values obtained at step 3 and divide it by the maximum measure obtained at step 4.

Finally, figure 3 shows co-occurrence between concepts (nodes) while taking into consideration their scientific domain/discipline.

4 Discussion

The following sections are dedicated to the discussion of the results obtained via text mining applied to the study of the literature on waste management. First, we'll discuss how connections between concepts and ERC domains/disciplines can help describe the scope of the existing waste management studies from 1959 till today. Second, by looking at co-occurrences and connections between various concepts in the entire sample, we'll show what current and future interdisciplinary research might look like. Finally, we'll discuss a well-known EU approach to waste management (i.e. "waste management hierarchy") and how it corresponds to the key concepts recurrent in the academic publications.

4.1 Field of waste management studies: cross-disciplinary conceptual connections

Based on the analysis of co-occurrences of key concepts both inside and in-between ERC domains/disciplines, we identified how recurrent waste management concepts connect one to another. The results are shown in a network graph on figure 3 below.

Figure 3: Concept Co-occurrence Network, per ERC domain/discipline

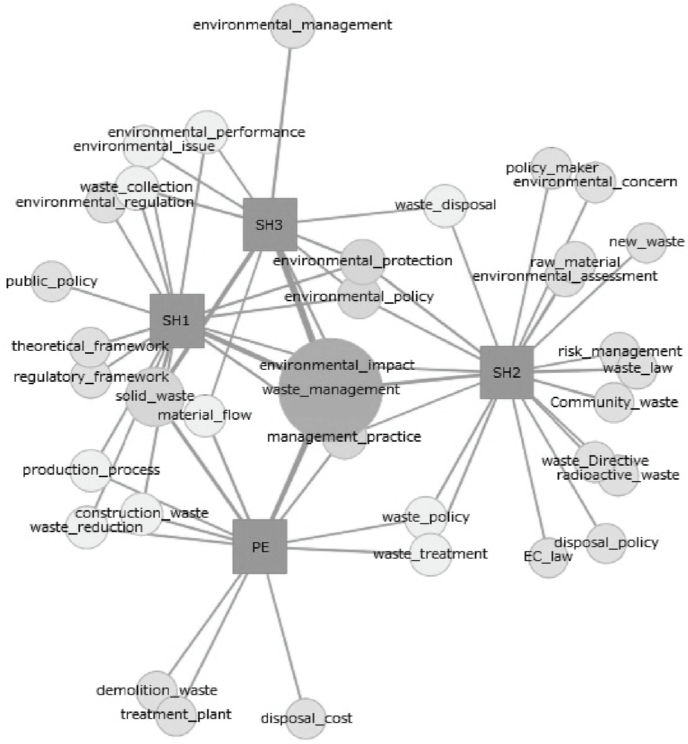


Figure 3 demonstrates a visualisation of connections between waste management concepts taking into consideration ERC domains/disciplines. The size of the nodes represents the frequency of concept use: larger circles are more frequently used concepts. Based on the reading of the results, we can distinguish between different categories of waste management concepts:

- ***General concepts*** – concepts common to all ERC domains/disciplines;
- ***Overarching concepts*** – concepts common to three out of four scientific disciplines;
- ***Common concepts*** – concepts used in at least two out of four scientific disciplines;
- ***Specific concepts*** – concepts specific to only one scientific discipline.

Classifying concepts according to their scientific affiliation could be used to create a tentative umbrella definition of waste management studies, common to all or most of the scientific disciplines involved in waste management research.

On of the applications of an 'umbrella definition' of waste management studies could be to underline the interdisciplinarity and/or fields of each discipline's expertise within waste management research. As a result, some directions for future research could follow. For instance, the topics that are currently studied by one scientific domain exclusively could in future be investigated from the perspective of other disciplines as well.

Based on figure 3, the most relevant interdisciplinary waste management subjects are represented by *general concepts* ("waste management" and "environmental impact") and *overarching concepts* (“environmental protection”, “environmental policy”, “solid waste”, and “management practice”). *Common concepts* (“waste policy”, “waste treatment”, “waste reduction”, “construction waste”, “production process”, “material flow”, “environmental performance”, “waste disposal”, “waste collection, “environmental issue”) result quite evenly distributed between different scientific domains. Finally, as for *specific concepts*, SH2 grouping, dominated by legal studies, has the highest concentration of concepts that belong to only one field of waste management research. Such studies seem to concentrate on normative and regulatory aspects of waste management, which limits their scope and applicability to other academic disciplines. Interestingly, the concept of “risk management” results as a specific SH2 concept, while it can be of potential interest for quantitative statistical studies, as well as business and management research. Similarly, subject of “disposal cost” (resulting as a specific PE concept) can be of great interest to SH1 economics and management studies and SH3 environmental studies research. Additionally, such expectedly ‘legal’ concepts as “public policy” and “environmental regulation” result in our sample as specific SH1 economics and management concepts.

Overall, figure 3 shows that PE domain concepts tend to talk about various stages of waste disposal (collection, treatment, disposal, disposal costs etc.). SH domain, on the other hand gravitates towards more general topics, such as management models and waste regulations. A systematic analysis of all the levels of concepts leads to the following broad description of waste management studies:
Waste management studies are generally focused on the investigation of the environmental impacts. Within this general context, social sciences give a peculiar emphasis on environmental protection and policies, whereas solid waste and management practices require integrated approach involving SH social studies (SH1 management disciplines in particular) and quantitative PE disciplines.

4.2 Concept networks and research streams

Figure 4 represent the concept network built on betweenness centrality principle and visualised without taking into consideration scientific domains. The dimensions of the circles in this case were standardized in order to ease its readability. The groups formed by the concepts connections can further help identify opportunities for future waste management research.

Figure 4: Waste Management Concept Network

Figure 4 shows a number of research streams within waste management studies, graphically seen as concepts' constellations. Some streams result richer and denser, particularly those oriented at research of a clear and shared definition of waste from a legislative standpoint (constellation 1 and 2 at the bottom left of the graph). Similarly, a stream of institutional and political economy research (constellation 3 at the top of the graph) is geared towards normative definitions, yet it appears slightly more interdisciplinary than strictly legal streams (1 and 2).

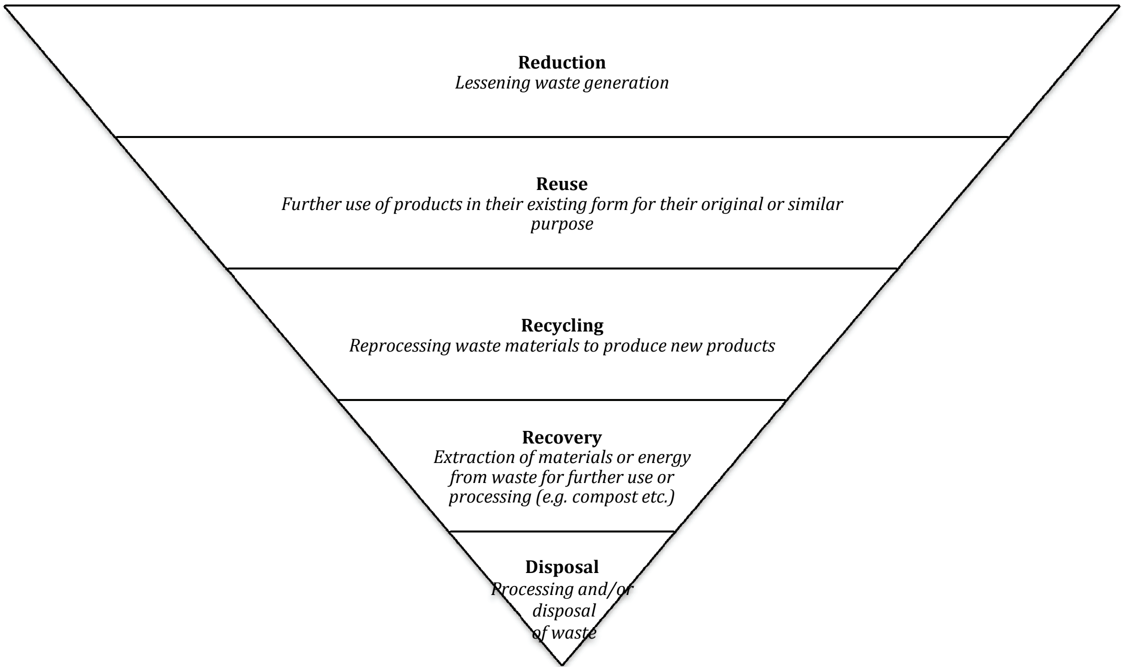
Another evident research stream (constellation 4 at the right) concerns managerial models, quality and performance. Constellations 5 (central part of the graph, slightly towards the bottom) and 6 (top left) seem to revolve around materials systems and processes, including evaluation of relative costs and investments efficacy.

Last but not the least are more 'technical' research streams that regard for instance: green product management and integrated production systems that control and minimize waste (constellation 7 on the top right); materials and product lifecycles (constellation 8 at the centre, slightly to the right); evaluation of energy costs and materials savings (constellation 9 at the left).

4.3 Academic research and EU policies: hierarchy and priorities

Waste management, as a high priority practical concern, is the subject of attention both by academics and policy makers. From European Union perspective, the recommended approach to waste management is summarized under 2008/98/EC Waste Framework Directive (EU Commission 2008), widely referred to as Waste Hierarchy (see figure 5). Such hierarchy aims to summarize the fundamental principles of waste management and to define the priority scale for possible alternatives in waste management decision-making. It's objective, naturally, is to promote and maintain sustainable production and process management on all levels.

Figure 5: EU Waste Management Hierarchy.



Adapted from 2008/98/EC Waste Framework Directive (EU Commission 2008).

Integrated management approach underlies EU Waste Hierarchy philosophy. According to the hierarchy, the top priority of efficient waste management is reduction and ideally complete avoidance of waste, while waste disposal is considered the least preferable option. Such approach treats waste as a possible resource rather than an unfortunate by-product of any human activity.

Comparing the waste hierarchy with the results obtained from the analysis of waste management literature in our sample leads to some meaningful considerations. As we can see on figure 3, one of the key general concepts of waste management literature in our sample is "environmental impact", followed by a number of subjects that are directly or indirectly linked to waste collection and waste disposal – the last option of the hierarchy. Thus, the scientific studies seem to consider the policies regarding materials' end of life rather than the management logic of waste minimization.

Such considerations are coherent with other studies (Price and Joseph 2000; Barrett and Lawlor 1997; Tjell 2005; Wilson 1996; Wilson 2007) that claim that waste hierarchy is not particularly efficient in driving governmental policies, and neither academic research. The complexity linked to waste disposal may be the reason why there is a considerable body of studies regarding the lower options of the waste hierarchy. For example, municipal solid waste management, one of the most challenging types of waste to move up the hierarchy, is the subject of numerous studies due to the fact that waste separation and collection procedures are critical to various stakeholders.

The text mining approach to the literature review in our sample seems to confirm the hypothesis that minimization and reduction waste management options occupy less space in the current waste management research than waste disposal and recovery studies. Consequently, our literature research demonstrates that existing waste management studies go somewhat contrary to the priorities defined by the EU waste hierarchy.

5 Conclusions

Academic literature on waste management has grown exponentially in the past years. While different scientific domains and disciplines actively engage in the waste management studies, there seems to be no common ground established that would allow a comprehensive definition of general waste management studies, their scope and limits.

Based on the lack of a sound and shared working definition, the object of this research is to analyse the international academic literature on waste management in order to identify the most relevant concepts and topics that currently occupy the attention of waste management scientists, analyse differences and similarities, as well as uncover directions for future research.

This literature review analysis was conducted via text mining technique applied to a representative sample of international scientific publications from 1956 till 2014. This analysis led to identification of distinctive words and recurrent concepts employed by various scientific domains and disciplines (ERC) involved into waste management research.

The results helped build taxonomy of relevant waste management concepts based on the connections between words and word combinations in various scientific disciplines. Specifically, general, overarching, common, and specific concepts were identified leading to creation of a tentative umbrella-definition of waste management studies: *Waste management studies are generally focused on the investigation of the environmental impacts. Within this general context, social sciences give a peculiar emphasis on environmental protection and policies, whereas solid waste and management practices require integrated approach involving SH social studies (SH1 management disciplines in particular) and quantitative PE disciplines.*

A further concepts analysis shows that some topics are studied exclusively by one scientific discipline, while it could benefit from more interdisciplinary perspectives. For instance, the concept of "risk management (currently, a concept specific to SH2 discipline) and "disposal cost" (specific to PE domain).

Additionally, the results show the most significant existing research streams, which were identified through statistical betweenness measures and visualised as 'constellations' inside concept networks. The existing studies seem to concentrate on normative research, institutional and political economy research, managerial models, and 'technical' (materials, energy and process management) research.
Finally, the research findings were contrasted with EU waste hierarchy, highlighting some critical points between normative directives and researchers' attention to certain topics.

The chosen methodology of text mining however has some limitations, which could have influenced the results. First, the concepts were identified as combinations of two words. Even though two-word concepts could be considered dominant in the scientific writings, future research may benefit from additionally analysing some one-word and three-and-more-word concepts. Furthermore, future research could be improved by refining the concept creation step and limiting insignificant or too general results. Alternatively, the automatic data analysis could be supplemented with a qualitative study of the concepts. Finally, future studies could execute text mining on a larger corpus of texts (e.g. full texts), which would require overcoming some considerable technical issues.

References

Barrett, Alan, and John Lawlor. 1997. 'Questioning the waste hierarchy: the case of a region with a low population density.' *Journal of environmental planning and management* 40 (1): 19-36.

Bolasco, Sergio. 2005. 'Statistica testuale e text mining: alcuni paradigmi applicativi.' *Quaderni di Statistica* 7.

Bolasco, Sergio, and Alessio Canzonetti. 2003. 'Sguardi sull'evoluzione dell'italiano standard degli anni Novanta, grazie al Text Mining e alla categorizzazione automatica del lessico del quotidiano "La Repubblica".' *Book of Short papers CLADAG-2003*: 57-60.

Calabrese, Giuseppe, and Deborah Morriello. 2014. 'Brand management come processo sociale. Un'indagine esplorativa sull'impatto dei nuovi internet brand touch-points.' Paper presented at the International Marketing Trends Conference 2014, Venice, Italy, January 24-25.

Cooper, Harris M. 1988. 'Organizing knowledge syntheses: A taxonomy of literature reviews.' *Knowledge in Society* 1 (1): 104-126.

Dulli, Susi, Paola Polpettini, and Massimiliano Trotta, eds. 2004. *Text mining: teoria e applicazioni*. Vol. 19. FrancoAngeli.

Egghe, Leo, and Christine Michel. 2002. 'Strong similarity measures for ordered sets of documents in information retrieval.' *Information processing & management* 38 (6): 823-848.

EU Commission. 2008. 'Directive 2008/98/EC of the European Parliament and of the Council of 19 November 2008 on waste and repealing certain directives (Waste framework directive, R1 formula in footnote of attachment II): http://eur-lex. europa. eu/LexUriServ.'

Freeman, Linton C. 1977. 'A set of measures of centrality based on betweenness." *Sociometry*: 35-41.

Fruchterman, Thomas MJ, and Edward M. Reingold. 1991.' Graph drawing by force-directed placement." *Softw., Pract. Exper.* 21 (11): 1129-1164.

Guérin-Pace, France. 1998. 'Textual statistics. An exploratory tool for the social sciences.' *Population* 10, no. 1: 73-95.

Jaccard, Paul. 1900. *Contribution au Probleme de L'Immigration Post-Glaciere de la Flore Alpine: Etude comporative de la flore alpine du massif du Wildhorn. du haut bassin du Trient et de la haute vallée de Bagnes*. Corbaz et Cie. "Bull. Soc. vaudoise Sci. nat.", 36: 87 – 130

Hansen, Mark Victor. 2015. 'Mark Victor Hansen | Mark Victor Hansen.' Accessed August 28. http://markvictorhansen.com/.

Huang, Anna. 2008. 'Similarity measures for text document clustering.' In *Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008), Christchurch, New Zealand*: 49-56.

Leopold, Edda, and Jörg Kindermann. 2002. 'Text categorization with support vector machines. How to represent texts in input space?.' *Machine Learning* 46 (1-3): 423-444.

Miner, Gary, John Elder IV, Andrew Fast, Thomas Hill, Robert Nisbet, and Dursun Delen. 2012. *Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications*. Academic Press.

Özgür, Arzucan, Burak Cetin, and Haluk Bingol. 2008. 'Co-occurrence network of reuters news.' *International Journal of Modern Physics,* 19 (05): 689-702.

Price, Jane L., and Jeremy B. Joseph. 2000. 'Demand management-a basis for waste policy: a critical review of the applicability of the waste hierarchy in terms of achieving sustainable waste management.' *Sustainable Development* 8 (2): 96.

Randolph, Justus J. 2009. 'A guide to writing the dissertation literature review.' *Practical Assessment, Research & Evaluation* 14(13): 1-13.

Tan, Ah-Hwee. 1999. 'Text mining: The state of the art and the challenges.' In *Proceedings of the PAKDD 1999 Workshop on Knowledge Disocovery from Advanced Databases*, vol. 8: 65.

Tjell, Jens Christian. 2005. 'Is the'waste hierarchy'sustainable?.' *Waste management and research* 23: 173-174.

Wilson, David B. 2009. 'Systematic coding.' *The handbook of research synthesis and meta-analysis* 2: 159-176.

Wilson, David C. 2007. 'Development drivers for waste management.' *Waste Management & Research* 25 (3): 198-207.

Wilson, David C. 1996. 'Stick or carrot?: The use of policy measures to move waste management up the hierarchy.' *Waste Management & Research* 14 (4): 385-398.