



Science Arts & Métiers (SAM)

is an open access repository that collects the work of Arts et Métiers Institute of Technology researchers and makes it freely available over the web where possible.

This is an author-deposited version published in: <https://sam.ensam.eu>
Handle ID: <http://hdl.handle.net/10985/19217>

To cite this version :

Louis RIVEST, Christian BRAESCH, Felix NYFFENEGGER, Christophe DANJOU, Nicolas MARANZANA, Frederic SEGONDS - Identifying PLM themes and clusters from a decade of research literature - International Journal of Product Lifecycle Management - Vol. 12, n°2, p.81-106 - 2019

Any correspondence concerning this service should be sent to the repository

Administrator : scienceouverte@ensam.eu



Identifying PLM themes and clusters from a decade of research literature

Louis Rivest*

Systems Engineering Department,
ETS,
Montreal, Quebec, Canada
Email: louis.rivest@etsmtl.ca
*Corresponding author

Christian Braesch

SYMME,
Université Savoie Mont Blanc,
Annecy, France
Email: christian.braesch@univ-smb.fr

Felix Nyffenegger

University of Applied Science Rapperswil (HSR),
864, Rapperswil, Switzerland
Email: felix.nyffenegger@hsr.ch

Christophe Danjou

Polytechnique Montréal,
Quebec, H3T-1J4, Canada
Email: christophe.danjou@polymtl.ca

Nicolas Maranzana and Frédéric Segonds

LCPI,
Arts et Metiers Institute of Technology,
HESAM Université,
Paris, France
Email: nicolas.maranzana@ensam.eu
Email: frederic.segonds@ensam.eu

Abstract: Product lifecycle management (PLM) encompasses a wide array of expertise, from designing green products to knowledge representation techniques. This paper characterises PLM as a research domain through the themes and clusters of a decade of scientific literature. Authors' keywords from 1,390 research papers published from 2005 to 2015 are analysed. The co-occurrence of these 2,947 normalised authors' keywords, connected in pairs via 11,289 edges, indicates how PLM research themes relate to each other to form communities – or clusters. These communities are revealed by filtering

the network according to the weights of the network's edges. The PLM core cluster, the PLM global cluster and the PLM overall cluster are distinguished based on the level of filtering, thus unveiling increasing levels of detail. The four major communities composing the PLM global cluster are 'interoperability', 'ontology', 'product data management' and 'lifecycle assessment'. The PLM overall cluster also reveals the 'intelligent product' community, which relates to the Industry 4.0 phenomenon. The BIM community is revealed as well, but remains isolated from the PLM overall cluster.

Keywords: bibliometric analysis; scientometry; product lifecycle management; PLM; review; mapping study; research communities; overview; literature review; systematic literature review; SLR.

Reference to this paper should be made as follows: Rivest, L., Braesch, C., Nyffenegger, F., Danjou, C., Maranzana, N. and Segonds, F. (2019) 'Identifying PLM themes and clusters from a decade of research literature', *Int. J. Product Lifecycle Management*, Vol. 12, No. 2, pp.81-106.

Biographical notes: Louis Rivest became a Professor at the Ecole de Technologie Supérieure in Montreal, Canada, after spending a few years in the aerospace industry. He obtained his PhD from the Ecole Polytechnique de Montreal, in 1993, and Bachelor in Mechanical Engineering, in 1988. His research centres on the models, methods, tools and processes supporting product development. His teaching and research activities thus relate to CAD and PLM.

Christian Braesch is an Associate Professor at the Université de Savoie and member of the System and Materials for Mechatronics (SYMME) Laboratory. His research and teaching activities are focused on enterprise modelling and more specifically on business process management and maturity models in PLM context.

Felix Nyffenegger is a Professor at the University of Applied Science Rapperswil. He obtained his PhD at the ETH Zurich in the field of the Fuzzy Front-end of Product Innovation, in 2006. Earlier, he studied Mechanical Engineering at the ETH. After his PhD, he spent several years in professional vendor-neutral PLM consulting. His research interest is focussed on methods and tools to organise product data along the complete lifecycle of a product type and its instances.

Christophe Danjou is an Assistant Professor at the Polytechnique Montreal. His research interests focus on interoperability between product development phases and Industry 4.0 issues for production engineering processes.

Nicolas Maranzana is an Associate Professor at the Arts et Metiers Institute of Technology in Paris, France and member of Product Design and Innovation Laboratory (LCPI). His research interests focus on the early stages of design process with the objective of developing new methods and tools to support designers' work.

Frédéric Segonds is an Associate Professor at the Arts et Metiers Institute of Technology and member of Product Design and Innovation Laboratory (LCPI). His research interests focus on the early stages of design collaboration, optimisation and collaborative design. This area includes the integration of stakeholders' core competences into the early stages of design, and providing methodologies and tools to support early product design.

This paper is a revised and expanded version of a paper entitled ‘Identifying PLM themes, trends and clusters through ten years of scientific publications’ presented at the PLM2016 Conference, Columbia, South Carolina, USA, 11–13 July 2016.

1 Introduction

Product lifecycle management (PLM) has gained momentum over the last two decades, both in industry and in the research community. PLM encompasses a vast array of topics, from designing green products to knowledge representation techniques. This variety makes it challenging to define the contour of the very nature of PLM.

This paper therefore presents a mapping of the PLM scientific literature published from 2005 to 2015 in peer-reviewed conferences and journals. It focuses on assessing a decade of PLM-related research work so as to establish a portrait of what themes and topics researchers have been working on over the decade. This paper specifically examines the clusters of themes and topics that are connected within PLM by studying the relationships between keywords found in the PLM literature. It offers a common baseline and vision to the PLM community that could promote transversal work.

Section 2 explains why the mapping study approach was chosen. Section 3 reviews existing PLM state-of-the-art literature. The methodology used for paper selection and the analysis technique are presented in Section 4. We present our findings in Section 5, followed by a discussion in Section 6. Section 7 contains our conclusions and some suggestions for future work.

2 Systematic literature review and mapping study

When a research area matures, there is often a substantial increase in the volume of publications and it becomes useful to classify these publications so as to provide a global overview of the domain to the academic community. This can be done by a systematic literature review (SLR), and more precisely, by a mapping study.

Out of the many definitions (Mulrow, 1994; Fink, 2013), we have chosen that proposed by Kitchenham et al. (2009): “A systematic literature review is a means of evaluating and interpreting all available research relevant to a particular research question, topic area, or phenomenon of interest. Systematic reviews aim to present a fair evaluation of a research topic by using a trustworthy, rigorous, and auditable methodology.” Such a review must be reproducible by other people following the same methodology. Furthermore, in this definition, a SLR aims on one hand to identify all the studies containing relevant elements to answer a research question, and on the other hand, to synthesise their outcomes. For these reasons, SLRs are often called *secondary studies*, while the studies forming their inputs are called *primary studies*. The increasing number of secondary studies has led to the emergence of new reviews, called *tertiary studies*, aiming at synthesising the secondary studies referring to the same research questions (Budgen et al., 2018).

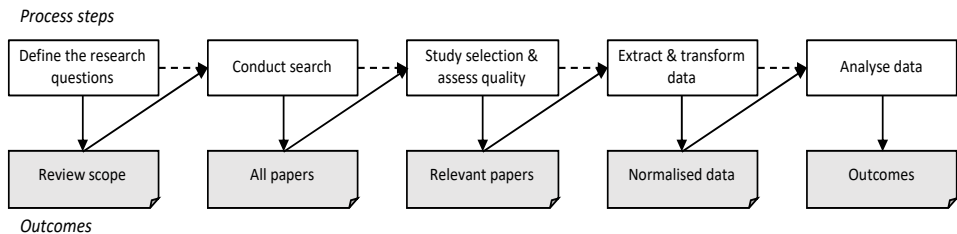
Many authors have proposed processes for conducting SLRs (Fink, 2013; Hart, 2018; Kitchenham et al., 2009; Okoli and Schabram, 2010). Some authors propose different classifications of SLRs (Grant and Booth, 2009; Kim et al., 2018). We focus here on distinguishing systematic maps and reviews (Kitchenham, et al., 2011; Petersen et al., 2015). A standard SLR aims to answer a very specific research question by aggregating the consistent and contradictory outcomes of primary studies. In contrast, a mapping study reviews a broader topic, classifies the relevant literature and aggregates studies with different categories.

In this work, we propose a mapping study to provide an overview of the scope of PLM and to highlight the connections between topics being researched, hence forming communities of topics within PLM as a research domain. The process we used to conduct this study is a synthesis of different mapping processes (Petersen et al., 2008, 2015). This synthesised process is composed of five essential steps:

- Define research questions: In a mapping study, these questions often identify the observations to be done to build the expected overview.
- Conduct a search: This step identifies on the one hand the scientific databases where the papers will be searched, and on the other hand, the strings used in the search engine. The outcome of this step is a set of primary studies extracted from different databases.
- Study selection and assess the quality of the primary study: Different inclusion and exclusion criteria are used to include or exclude studies that are relevant – or not – for the review.
- Extract and transform data: Extract the relevant data for the study and transform it into a consistent form.
- Analyse data: This step performs the different analyses to provide the expected outcomes.

Figure 1 summarises this process.

Figure 1 Mapping study steps



3 PLM literature review or state-of-the-art summaries

In this section, we briefly review existing works dealing with PLM literature reviews or state-of-the-art summaries. It can be noted that very few secondary studies of the PLM domain have been published to date.

In 2005, Ming et al. presented a PLM review, which now dates back to over a decade. More than five years later, Cao and Folan (2012) reviewed the evolution of product lifecycle literature from 1950 to 2009, showing how it led to PLM. In 2013, Nappi and Rozenfeld (2013) published a PLM literature review focusing on sustainability performance indicators. Several reviews were completed in 2015, including Bhatt et al. (2015), who published a PLM review identifying some of the most frequent PLM research themes published in PLM international conferences. Varandas et al. (2015) performed an analysis of PLM literature published from 2006 to 2010, with a focus on the concepts of PLM, new product development (NPD), environmental sustainability and their interfaces (in Portuguese). Mas et al. (2015) published a review of PLM from the perspective of its impact on the aerospace industry, in which they presented selected PLM topics from the developments in aerospace over roughly the last 50 years. Some papers, such as those by Laumond (2006) and Pulkkinen et al. (2013) have promising titles but their contents do not help us in our work. The classical paper by Terzi et al. (2010) shares some of our aims and describes the constituent elements of PLM, classifying them into three fundamental themes: information and communication technologies, business processes and methodologies. Corallo et al. (2013) propose a critical review of PLM definitions and describe the PLM landscape. Their description is organised around the concepts of people, technologies and processes. While they further describe six aspects of PLM, from the ‘strategic business approach’ to collaboration, their description of the PLM landscape does not actually describe the details of these six aspects.

David and Rowe (2015) consider that academic research on PLM is mainly focussed on technical issues and instead propose research avenues focusing on the organisational and managerial issues of PLM systems. More recently, Holler (2018) mapped the field of PLM based on techniques such as citation analysis, and interestingly, authors’ keywords analysis, yielding a graph of 40 connected PLM keywords, but these were not organised into communities.

None of the above-mentioned papers exposes global PLM trends and clusters over the years to produce a global map of PLM themes. This paper proposes a PLM mapping study based on authors’ keywords analysis. Hence, this paper offers a unique, evidence-based perspective on PLM communities of topics.

4 Method used to conduct our mapping study of the PLM domain

In this section, we describe how we applied the above-described five-step mapping study to achieve our PLM domain analysis.

4.1 Define the research questions

The goal of our mapping study is to draw a general portrait of PLM scientific literature over the last decade. This leads to the following two research questions:

RQ1 What are the major themes being studied in the PLM scientific literature? This question aims to reveal what researchers consider to be part of PLM science.

RQ2 What are the clusters of ideas that are logically connected within the PLM domain? These clusters are also designated as ‘communities’ of ideas or keywords.

4.2 Conduct the search

In this section, we present the search protocol used to collect all the relevant papers. The objective is to identify the databases where the papers have to be searched as well as the criteria to be used with search engines to identify the appropriate papers.

We searched papers from the following four sources:

- *International Journal of Product Lifecycle Management (IJPLM)*
- Product Lifecycle Management International Conference (PLM IC)
- Web of Science
- Science Direct.¹

As the context of our study is scientific literature, we want to identify papers with an acceptable scientific standard, and so we choose to retain peer-reviewed conference and journal papers published during the last decade. Searching scientific databases requires a set of criteria to be defined with which to identify the papers relevant for our study. As the scope of our study is PLM, the first criteria are that:

- each paper must have PLM (or any alternative spelling such as product life cycle management) in the title, keywords, abstract, or title of the proceedings or journal
- each paper must have a date and must have been published between 1st January 2005 and 31st December 2015
- each paper must be published in a peer-reviewed conference or journal.

The first criterion implies that all papers included in our work could be said to be ‘PLM-tagged research’. Hence, papers that could have otherwise been classified as belonging to PLM because of their content, such as, for example, ‘collaboration and interoperability in product development’, were not included if they were not PLM-tagged by their authors. As a consequence, all publications included in our analysis are considered as PLM-related by their authors. This is consistent with our intent to delineate the nature of PLM research. It should be emphasised that all the papers from the *IJPLM* and PLM IC satisfy this criterion, even when a paper does not use PLM in its title or keywords.

In our work, we consider that the main ideas of a paper are identified in its list of keywords. As we want to highlight the links between these keywords, we also retain the following criteria:

- the paper must be written in English
- the paper must have keywords.²

All papers retained by applying these criteria were stored in a Zotero database. During this search, we used a Zotero connector linked to our web browser to automatically capture the meta-information of the various papers.

During this step, we also collected PDF files (whenever possible) linked to the individual papers; these files are used in the next step of our process. At the end of this step, we have a Zotero database containing papers meeting the set of specified criteria.

4.3 Study selection and assess quality

Before reviewing all the papers stored in the Zotero database, it is necessary to include all the papers missed by the previous step. Some authors argue for the use of backward snowball sampling (Webster and Watson, 2002; Wohlin, 2014; Jalali and Wohlin, 2012). This method can be summarised in three steps:

- 1 Identify a starting set of papers from which to go backward.
- 2 Go backwards by reviewing the citations for the papers identified in Step 1 to determine prior articles to be considered (iterate until no new papers are identified).
- 3 Go forward by identifying papers citing those that were identified in the previous steps.

In this study, we chose to use four major sources of publications. All papers from the PLM IC and the *IJPLM* were included in the first step. Moreover, all of the PLM-tagged papers proposed by the Web of Science and Science Direct databases were also obtained in the first step. Having included all of those, we consider that it is not necessary to apply the snowball sampling method to find new papers.

As there are no new papers, the next phase of this step is to check the data recorded in the Zotero database to improve our data quality. This is done by comparing the PDF file of each publication with the data stored in the Zotero database. We assume that the reference is the PDF file; if a difference is found between a PDF file and the Zotero data, we correct the latter.

During this examination, we also applied different criteria to exclude the papers that are not relevant for our study. Above all, duplicate publications should not exist in the Zotero database. Next, we verified that the retained conference and journal papers were published after a peer-review process to ensure their scientific quality. This was done by looking for a clear indication of a peer-review process on each conference and journal website. As the PLM IC and the *IJPLM* follow this process to validate all of their publications, we only had to check the papers proposed by Web of Science and Science Direct. Moreover, if PLM is used in the title, keywords or abstract, or in the name of the proceedings or of the journal, this acronym must stand for 'product lifecycle management'. Lastly, we consider that the authors are best qualified to explain their own work based on keywords; thus, this is an adequate starting point to analyse what topics are being researched and how they are linked to each other. Hence, keywords added by editors were removed for our study, since our approach is based on authors' keywords.

The following criteria were therefore used to exclude publications from the database:

- A publication is duplicated in the database.
- The conference or the journal does not follow a peer-review process.
- The acronym PLM does not stand for 'product lifecycle management'.
- The publication was not provided with at least one keyword by the author(s).

At the end of this step, each publication retained is qualified to respect our filtering criteria and, therefore, to meet acceptable scientific standards.

4.4 *Extract and transform the data*

First, data was exported into a specific database in order to be processed. In this database, the *publications* table stores all the required data of the different primary studies, while the *tags* table identifies all the keywords in a primary study. The *publicationsTags* table links each keyword to the publications.

This database was updated automatically by an algorithm developed for that purpose. The algorithm follows the rules defined by the following phases:

- Field alignment: Fields from different sources need to be aligned because they have the same meaning; for example, the journal's name or the name of a conference.
- Keyword normalisation: Each keyword is translated into a normalised form, so that keywords with the same meaning are unique.
- Acronym transformation: The same acronym may have different meanings, and so the correct meaning must be searched for in the abstract or in the PDF file; acronyms are replaced by their definitions.

The automation of this process allows the full process to be repeated in the future, so as to include new publications and keep our database up to date. The most critical step is the normalisation. The journal and proceedings' names, and most importantly, the keywords, had to be transformed into a normalised form.

The keyword normalisation process is as explained in Nyffenegger et al. (2016). This process includes steps such as:

- American English is chosen over Britain English (e.g., 'modelling' is normalised to 'modeling', 'visualisation' to 'visualization', etc.).
- Plurals become singular except when a word is usually used in plural, such as 'logistics'.
- Lifecycle is always written as one word.

4.5 *Analyse data*

Co-occurrence networks are used to visualise the connections between concepts expressed by terms in text sources. In this context, co-occurrence is defined as the paired presence of two terms in a specified unit of text. Two co-occurring terms can be called 'neighbours', which are grouped into 'neighbourhoods' based on their interconnections.

In this study, two keywords are neighbours if they appear in the same publication. A co-occurrence network then helps to visualise the relationships between keywords and highlights clusters of keywords by identifying their neighbourhood. In our keyword co-occurrence network, each keyword is represented by a node. The importance of a keyword is characterised by its count. The count corresponds to the number of occurrences of a keyword – or a node – amongst all the retained publications.

Co-occurrence relationships are represented as edges; hence the number of co-occurrences of a given pair of keywords influences the weight of that edge.

McSweeney (2009) and Heymann (2006) discuss different ways to measure the importance of a node in a graph. One appropriate measure is the degree. The degree corresponds to the number of edges that are attached to a node. A node – or keyword – with a high degree plays a central role in the network.

One fundamental interest in network analysis is the detection of communities. A community can be described as a group of densely connected nodes that has many edges between the members of a group, but only a few edges between the different groups. In our study, we aim to identify the core topics of PLM based on research publications, as well as to characterise communities in this research domain.

Newman (2006) discusses different approaches for analysing the topology of a network graph, such as graph partitioning and community structure methods. Community structure methods assume that the network of interest divides naturally into subgroups. If the experimenter's job is to find those groups, he particularly suggests using optimal modularity, referring to the work of Guimera and Amaral (2005) and Danon et al. (2005). The idea behind this approach is to quantify the assumption that the network of interest has a higher level of structure than that of a random network. The corresponding measure is called modularity. The modularity is the number of edges falling into groups minus the expected number in an equivalent network with edges placed at random. Using algorithms, the network can now be divided into groups such that the modularity is optimised. The algorithm used in this work to analyse the network is an implementation of optimal modularity according to Blondel et al. (2008).

5 Results and findings

The results are presented according to two different perspectives. Keywords are first analysed as a list of topics in order to assess their relative importance in the PLM domain. The connections between keywords are then used to obtain graphs that show the links or connections between these topics. Filtering the data based on the weights of these connections reveals distinct communities and thus a second perspective.

5.1 *PLM research themes as revealed by authors' keywords count*

There are 2,947 normalised authors' keywords extracted from the list of 1,390 retained publications. These 2,947 nodes are connected in pairs via 11,289 edges.

The bar chart in Figure 2 shows the distribution of the count of normalised authors' keywords. The top-ranking keyword, 'product lifecycle management', was removed for clarity, and the graph only includes keywords with a count of two or more, for a total of 745 normalised keywords. The 'count' corresponds to the number of times a keyword was found in a paper from our database. Hence, a keyword with a count of 10 was used in ten different papers.

The L-shape suggests that a large number of keywords have a limited influence in defining PLM as a research domain. More precisely, we find that only 72 normalised keywords have a count of 10 or more, which leaves 2,874 normalised keywords (97.5%) with a count of less than 10. In the list of 2,947 normalised keywords, 2,201 (74.7%) have a count of one. Including PLM, there are 18 authors' keywords with a count of 20 or more, as listed in Table 1.

Figure 2 Normalised authors' keywords count distribution (without PLM) (see online version for colours)

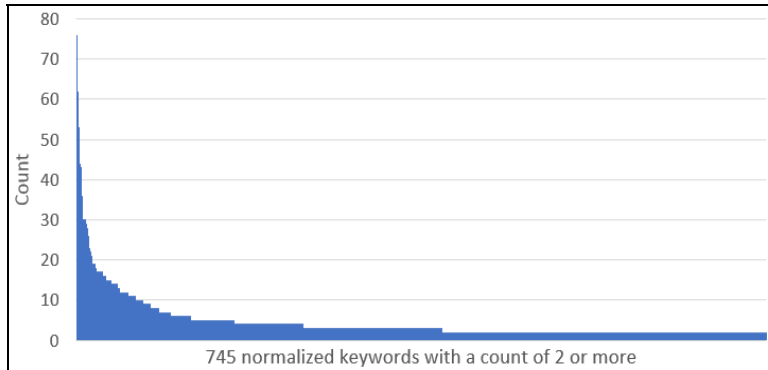


Table 1 Top-ranking normalised keywords by their count

	<i>Normalised keywords</i>	<i>Count</i>
1	product lifecycle management	642
2	knowledge management	76
3	ontology	62
4	product data management	53
5	product development	44
6	interoperability	43
7	product lifecycle	41
8	computer aided design	36
9	concurrent engineering	30
10	lifecycle assessment	30
11	new product development	30
12	collaborative design	29
13	collaboration	28
14	product lifecycle management system	26
15	sustainability	24
16	collaborative engineering	23
17	knowledge based engineering	22
18	service oriented architecture	21

Not surprisingly, ‘product lifecycle management’ has the overwhelmingly highest count, with 642. We should note that the keyword PLM (normalised as ‘product lifecycle management’) is not used by all papers; this can be deemed as reasonable for papers published via a medium dedicated to PLM (*IJPLM* and PLM International Conference) where PLM as a keyword may be considered implicit.

The second-ranked keyword in the list is ‘knowledge management’ with a count of 76. The keywords ‘ontology’, ‘product data management’, ‘product development’ and ‘interoperability’ appear next.

Our focus is on identifying the communities of keywords that compose PLM as a research domain. Hence, in the following pages filtering is performed based on the weights of the edges, rather than on the keywords' count or degree. The weight of an edge corresponds to the number of papers where a pair of keywords co-occurred. Hence, a weight of 5 means that the two keywords that it connects are found as a pair in five articles.

Table 2 Top-ranking co-occurrences (edges) of normalised keywords

	<i>Normalised keyword 1</i>	<i>Normalised keyword 2</i>	<i>Weight</i>
1	product lifecycle management	knowledge management	48
2	product lifecycle management	product data management	35
3	product lifecycle management	ontology	31
4	product lifecycle management	product development	26
5	product lifecycle management	interoperability	20
6	product lifecycle management	concurrent engineering	19
7	product lifecycle management	new product development	19
8	product lifecycle management	computer aided design	18
9	product lifecycle management	collaboration	16
10	product lifecycle management	knowledge based engineering	16
11	product lifecycle management	service oriented architecture	16
12	product lifecycle management	engineering change management	15
13	product lifecycle management	collaborative engineering	13
14	product lifecycle management	innovation	11
15	product lifecycle management	standard	11
16	product lifecycle management	supply chain management	11
17	product lifecycle management	XML	11
18	product lifecycle management	collaborative design	10

Figure 4 Distribution of edges' weight (see online version for colours)

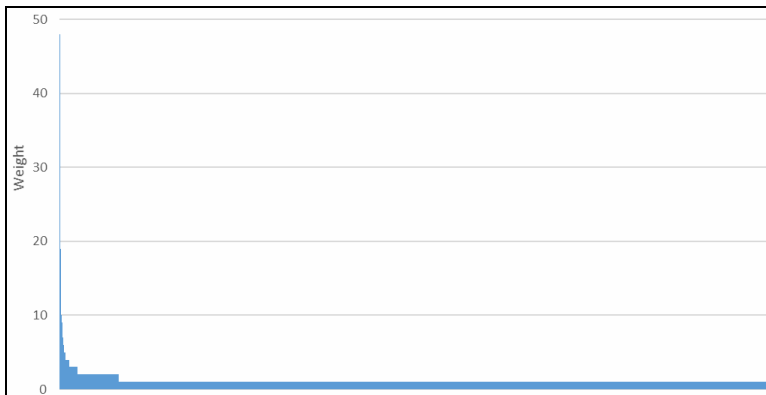


Table 2 shows the top-ranking co-occurrences of keywords, which corresponds to the edges with the highest weight on the graph. Only edges with a weight of 10 or more are

listed in Table 2. We find 18 edges that have a weight of 10 or more. We see that ‘product lifecycle management’ and ‘knowledge management’ appear jointly as keywords in 48 distinct papers, hence the edge that connects them has a weight of 48. We also observe that all the edges listed connect ‘product lifecycle management’ to another keyword. That is as expected, since we saw above that this keyword has overwhelmingly the highest count.

Table 3 Top-ranking co-occurrences (edges) of normalised keywords without PLM

	<i>Normalised keyword 1</i>	<i>Normalised keyword 2</i>	<i>Weight</i>
1	ontology	semantic web	8
2	computer aided design	product data management	7
3	ontology	knowledge management	7
4	ontology	OWL	7
5	interoperability	data exchange	6
6	interoperability	standard	6
7	lifecycle assessment	sustainability	6
8	interoperability	product data management	5
9	interoperability	product lifecycle management system	5
10	product structure	product data management	5
11	ontology	SWRL	5
12	building information modelling	building information management	4
13	intelligent product	closed loop product lifecycle management	4
14	interoperability	collaborative engineering	4
15	computer aided design	computer aided engineering	4
16	data dissemination	fuzzy analytic hierarchy process	4
17	interoperability	open standard	4
18	data exchange	product lifecycle	4
19	data exchange	standard	4
20	service oriented architecture	enterprise service bus	4
21	internet of things	intelligent product	4
22	interoperability	ontology	4
23	interoperability	product lifecycle support	4
24	knowledge management	new product development	4
25	lifecycle assessment	lifecycle costing	4
26	lifecycle assessment	sustainable development	4
27	OWL	reasoning	4
28	OWL	semantic web	4
29	product design	product lifecycle	4
30	service oriented architecture	standard	4
31	STEP	standard	4

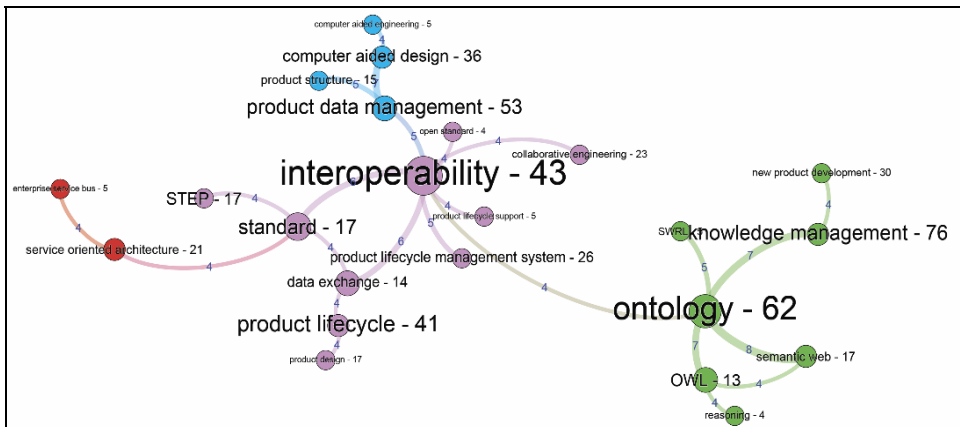
The graph in Figure 4 shows the distribution of the edges according to their weight. Just as was seen for the keyword count distribution (Figure 2), most of the edges have a very low weight. More precisely, out of a total of 11,289 edges, we see that 10,354 edges (91.7%) have a weight of 1. A total of 11,151 edges have a weight of 1, 2 or 3 (98.8%).

As indicated in Table 2, the top-ranking edges involve ‘product lifecycle management’ as a keyword, which leads to the ‘PLM spider graph’ in Figure 3. Hence, in order to reveal the communities, or sub-clusters, of PLM as a research domain, it is deemed useful to remove PLM as a keyword. This leaves a total of 9,832 edges that do not involve PLM as a keyword. Amongst those edges, 31 have a weight of 4 or more, as listed in Table 3, connecting 34 distinct keywords. We observe that ‘ontology’ and ‘semantic web’ are connected by an edge of weight 8. Various graphs, described in the following sections, are generated from this data, revealing distinct PLM communities.

5.2.1 PLM core cluster

Figure 5 were generated using Gephi v.0.9.1. The spatialisation algorithm used is Force Atlas 2. Details about this algorithm can be found in Jacomy et al. (2014). Gephi’s modularity class function is used to reveal communities. Modularity reveals the communities within a graph – or sub-clusters – as discussed in the methodology section. Different communities are displayed using different colours to help distinguish them visually.

Figure 5 PLM core cluster – revealed by the edges of weight of 4 or more (without PLM) (see online version for colours)



Filtering the data based on the weights of the edges in order to keep the edges (and associated nodes) that have a weight of four or more reveals the *PLM core cluster* (Figure 5). This graph size is a convenient starting point for analysis.

The PLM core cluster graph shown in Figure 5 has 23 nodes and 24 links. This core cluster is composed of four communities organised around ‘ontology’, ‘interoperability’, ‘product data management’ and ‘service oriented architecture’. The central node of the PLM core cluster is ‘interoperability’, as confirmed by its relatively high degree of 8, which translates into a bigger font size. This can be observed even if the count of

‘interoperability’, at 43, is lower than that of ‘product data management’, at 53. The second major node is ‘ontology’ with a degree of 5.

The *‘interoperability’ community*, displayed in violet, acts like a bridge to the other clusters in this graph. The keyword ‘interoperability’ has eight connections to eight first-level neighbours related with a weight of 4 or higher. We observe that ‘interoperability’ is connected to ‘standard’ and to ‘data exchange’ with a weight of 6, and to ‘open standard’ with a weight of 4. The keyword ‘standard’ is connected to ‘STEP’, which is not surprising since STEP is a widely used standard for PLM.

Moreover, interoperability supports collaborative engineering. We observe that ‘product lifecycle support’ and ‘product lifecycle management system’ have a first-order connection to ‘interoperability’. In the lower part of Figure 5, there is a connection between ‘data exchange’ and ‘product lifecycle’.

The ‘interoperability’ community is connected to the three other communities: ‘service oriented architecture’, in red, ‘product data management’, in blue, and ‘ontology’, in green.

The *‘ontology’ community*, in green, revolves around the keyword ‘ontology’ that counts connections to five first-level neighbours related by a weight of 4 or more (its degree is 5). The keyword ‘knowledge management’ (or KM) is a major keyword of the PLM core cluster, with a count of 76. This is the second most-used keyword, after ‘product lifecycle management’. Interestingly, even if KM is a top node, it is not the central node of the PLM core cluster.

The five first-level neighbours of ‘ontology’ are ‘semantic web’, ‘OWL’, ‘SWRL’, ‘knowledge management’ and ‘interoperability’. Since an ontology is a tool used to represent a body of knowledge (Noy and Klein, 2004), observing a direct link to knowledge management seems intuitive. Having ‘ontology’ connected to ‘SWRL’ also seems logical since semantic web rule language is a set of language rules saved as part of an ontology.

Along the same lines, Web Ontology Language (OWL) is a language used to represent and process information. Hence, having the keyword ‘reasoning’ directly connected to OWL is to be expected, since reasoning is purposely supported by OWL. Interestingly, KM is directly connected to ‘new product development’.

The *‘product data management’ community*, in blue, connects ‘product data management’, ‘product structure’, ‘computer aided design’ and ‘computer aided engineering’. These connections highlight the main concerns of PLM research in that community. The connection between ‘product data management’ (PDM) and ‘interoperability’ is particularly interesting, since PDM is not a tool for interoperability per se, while interoperability may well be a question of interest to the PDM community.

The *‘service oriented architecture’ community*, in red, is connected to the ‘interoperability’ community through the keyword ‘standard’. This community is composed of ‘service oriented architecture’ (SOA) and ‘enterprise service bus’.

Now that the PLM core cluster, revealed by filtering data based on keeping edges of a weight of 4 or more, has been described, we can add the edges that have a weight of 3 in order to reveal the PLM ‘global cluster’.

5.2.2 PLM global cluster

The graph shown in Figure 5 is designated as the ‘PLM core cluster’, since it shows the core content of PLM as a research domain. The graph shown in Figure 6 corresponds to

logic', 'UML' and 'web service', and thus a variety of tools, or languages, that are used to manipulate knowledge. We hence consider that this community has two sub-groups: the concepts associated with knowledge and the tools used to manipulate them.

The '*interoperability*' community, in violet, now has 18 nodes. We note the presence of two keywords about collaboration ('collaborative engineering' and 'collaborative design'), as well as 'product design'. This community also includes 'product model', 'integrated model' and 'industrial digital mockup'. On a more technical note, we find 'open standard', 'standard' and 'data exchange'. The interoperability community also contains 'product lifecycle' as well as 'product lifecycle management system' and 'product lifecycle support'. Interestingly, research in this community is focussed on technical aspects, while the social or cultural aspects in interoperability are absent. This specific data tends to agree with David and Rowe's (2015) statement about the fact that PLM research is focussed on technical aspects.

We also observe that 'STEP' is now included in the product data management community. Another change is that the 'service oriented architecture' that was a distinct community in Figure 5 (the core cluster) is now integrated into the 'interoperability' community in the global cluster.

Interestingly, 'lifecycle impact assessment' appears here as a keyword in the interoperability community, while it would have been expected to be part of the lifecycle assessment community, in grey.

The '*product data management*' (PDM) community, in blue, now counts 15 nodes. The three dominant nodes are 'product data management', 'computer aided design' and 'concurrent engineering'. This community includes 'product development', 'product development process' and 'engineering design'. It also includes the tools 'computer aided design', 'computer aided manufacturing' and 'computer aided engineering'. Interestingly, 'product data management' is linked to 'enterprise resource planning', along with 'design'. The keyword 'concurrent engineering' is connected to 'collaboration', 'product structure' and 'digital mockup'. The keyword 'virtual reality' is also part of this community.

We observe that the PDM community is linked to the 'interoperability' community through four links: 'product data management' is linked to 'interoperability', 'concurrent engineering' is connected to 'product lifecycle', and 'STEP' connects to 'standard' and to 'interoperability'. We can also observe that 'STEP', while belonging to the PDM community, is at the interface of both the 'interoperability' and 'ontology' communities.

We also notice that the 'interoperability' and the 'ontology' communities are intertwined through five connections: 'interoperability' is connected to 'OWL', 'ontology', 'semantic web' and 'reasoning', while 'standard' is connected to 'ontology'.

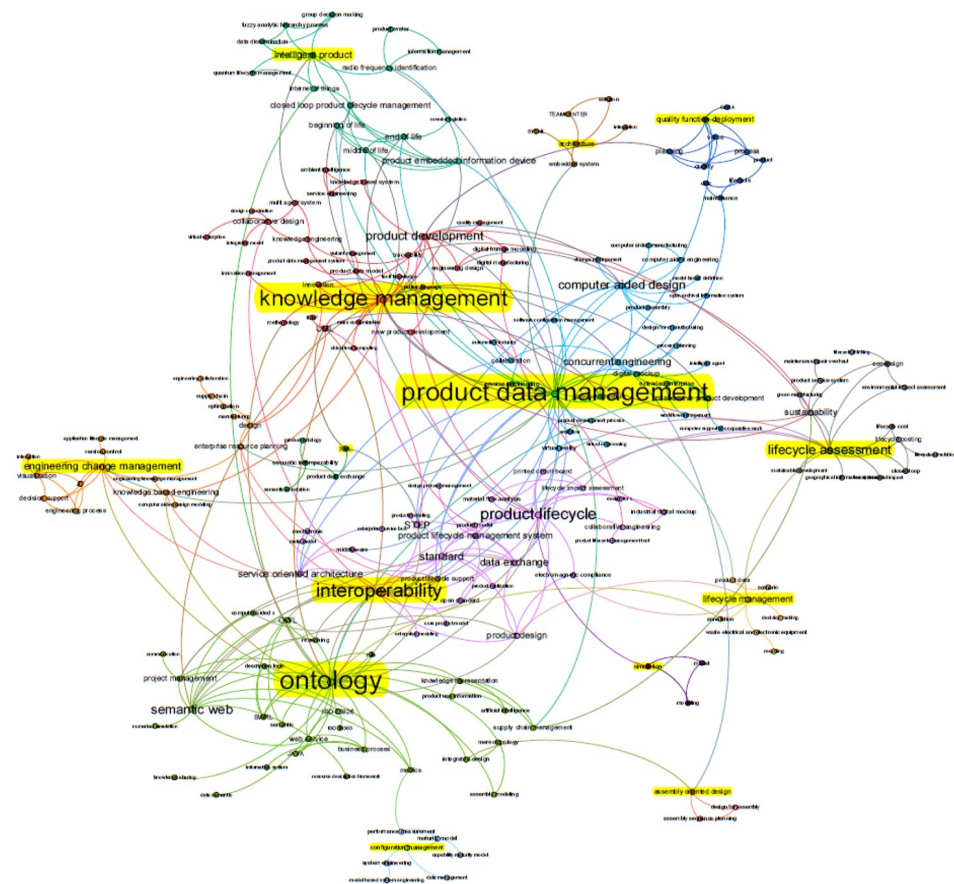
Overall, we note that the 'lifecycle assessment' community is one of the four major components of the PLM global cluster, even though it is only linked through one edge, while the three other communities are more intertwined, as stated above.

Finally, we should mention that other isolated communities are revealed when filtering the data based on edge weights of 3 or more. Amongst them are two especially interesting ones: the 'building information modeling' community and the 'intelligent product' community. However, these are even more interesting to analyse when revealed by filtering the data based on edges with weights of 2 or more, as discussed next.

5.2.3 PLM overall cluster

Thus far, we have presented the PLM core cluster (Figure 5) and the PLM global cluster (Figure 6). Both describe the same reality of PLM as a research domain, but through different levels of detail. They were presented in sequence so as to facilitate understanding of the data. Now that the reader is familiar with them, we add some more details to the description of PLM as a research domain by filtering the data to include edges that have a weight of 2 or more, thus revealing the *PLM overall cluster*. It is composed of 14 communities. The central keyword of each community is highlighted in Figure 7. Amongst these communities are ‘product data management’, ‘ontology’, ‘interoperability’ and ‘lifecycle assessment’, as in the global cluster; the ‘knowledge management’ community is now distinct from the ‘ontology’ one. Other new communities also appear: ‘intelligent product’, ‘configuration management’, ‘lifecycle management’ and ‘engineering change management’. Five smaller communities are also revealed: ‘assembly oriented design’, ‘simulation’, ‘quality function deployment’, ‘architecture’ and ‘XML’.

Figure 7 PLM overall cluster – revealed by the edges of weight of 2 or more (without PLM) (see online version for colours)



‘mechatronic’. Two nodes, ‘collaborative design’ and ‘integrated model’, moved to the knowledge management community.

Taken globally, the variations in the composition of PLM communities when filtering from weights of three or more to weights of two or more may indicate that the latter level of filtering is very sensitive, possibly too much so, to characterise communities in a stable manner. On the other hand, such sensitivity may help to identify upcoming trends. Incidentally, filtering on edge weights of two or more reveals two trending communities: ‘intelligent product’ and ‘building information modeling’, both described next.

The *intelligent product community* is shown in Figure 11. This community relates to the Industry 4.0 phenomenon, even if it is not announced as such. Indeed, keywords such as ‘product embedded information device’, ‘radio frequency identification’ and ‘internet of things’ refer to concepts or technologies being used by Industry 4.0 for ubiquitous connectivity and real-time reaction, as described by Danjou et al. (2017).

Figure 11 The *intelligent product community* from the PLM overall cluster (see online version for colours)

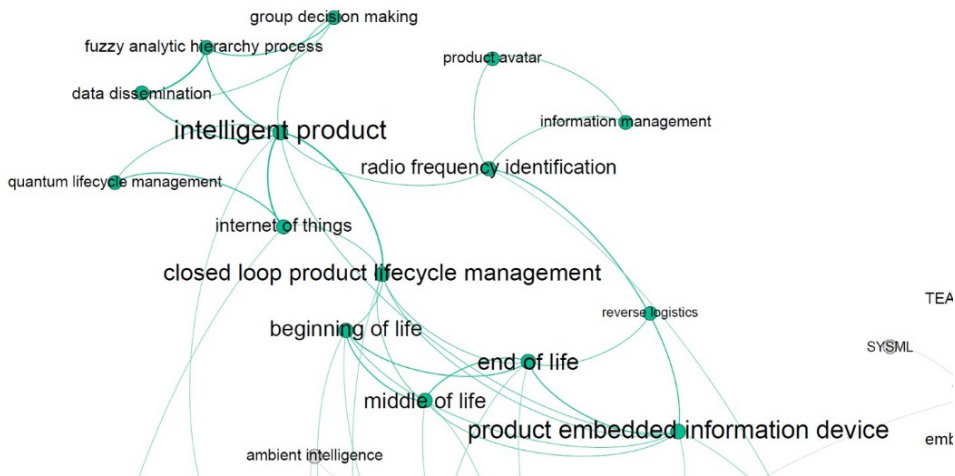
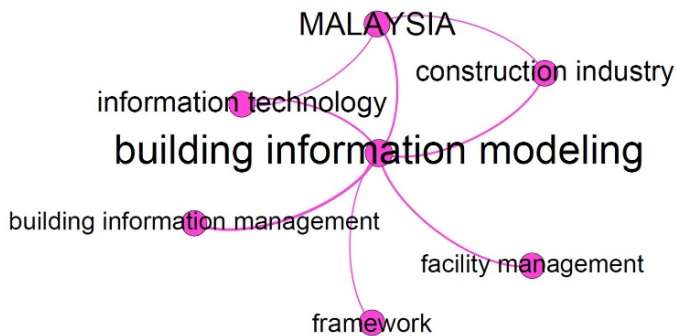


Figure 12 The *building information modeling* community from the PLM overall cluster (see online version for colours)



The *building information modeling (BIM) community* at this level of filtering (edges with weights of 2 or more) is shown in Figure 12, with its seven nodes. The first observation is that the BIM community remains completely isolated from the PLM overall cluster, and hence, is not shown in Figure 7. This could indicate that the BIM community of research is not yet integrated with PLM research. Next, we observe that ‘building information modeling’ and ‘building information management’ belong to this community. On the more technical side, we find ‘framework’ and ‘information technology’, while on the application side, we find ‘construction industry’ and ‘facility management’.

6 Discussion

6.1 On the use of keywords

In this study, we chose to focus on authors’ keywords. It is assumed that the authors are best qualified to summarise their work with their selection of keywords. Globally, keywords can be defined in different ways, such as by the authors, by a journal’s prescribed categories, or by automatic tags generated at the proof stage (Hartley and Kostoff, 2003). Some keywords can also be generated by users which is the case in some collaborative platforms such as CiteULike (Kipp, 2005). In such a context, the author keywords tend to achieve higher recall than collaborative tags (Lu and Kipp, 2014). In the end, as concluded by Hartley and Kostoff (2003), it seems appropriate to rely on the authors’ keywords when conducting a keywords-based SLR, since authors are best qualified to choose the keywords for their own work.

Assuming that using authors’ keywords is a good way to express the content of scientific papers, the use of keywords faces some limitations. On one hand, different keywords may be used by different authors to express a single idea. On the other hand, some keywords are much more specific than others. For example, ‘waste electrical and electronic equipment’ is a rather precise keyword, while ‘model’ is very generic. Hence, in future work, it may be worthwhile to add a normalisation step in order to harmonise the level of abstraction of the considered keywords – no trivial task.

In addition, the keyword normalisation rules used here did not lead to the merging of similar keywords. For example, in the lifecycle assessment community (Figure 6), we find two similar keywords: ‘lifecycle cost’ and ‘lifecycle costing’. They could have been merged if different normalisation rules had been used, which may have slightly modified the results.

6.2 On filtering levels

As mentioned above, filtering by including edges with weights of two or more is possibly too sensitive to yield stable results. Indeed, one or two newly-published papers with a given pair of keywords could possibly modify the communities. Hence, the most representative and stable portrait of PLM research is most likely the one obtained when filtering by edges with a weight of 3 or more, as in Figure 6 (the PLM global cluster).

It is also interesting to observe that some keywords are not present in the results. For example, some trendy keywords such as ‘4.0’ or ‘cyber physical system’ (CPS) are not present at all in the unveiled PLM research community.

6.3 *Answering the research questions*

Looking back at the research questions, we should now be able to provide answers based on the data presented.

RQ1 What are the major themes being studied in the PLM scientific literature?

The answer to this first question lies in the list of major keywords presented in Table 1. The five major keywords by count are (except for ‘product lifecycle management’): ‘knowledge management’, ‘ontology’, ‘product data management’, ‘product development’ and ‘interoperability’. However, the results also show that while 745 keywords with a count of two or more were found within 1,390 publications, only 18 keywords have a count of 20 or more, while a count of 20 means that the keyword appears in 1.4% of the considered papers. The ‘strongest’ of these top 18 keywords, ‘knowledge management’, with a count of 76, thus appears in 5.5% of papers. Hence, PLM as a research domain has a coverage that is both wide and thin.

RQ2 What are the clusters of ideas (or ‘communities’) that are logically connected within the PLM domain?

The four major communities shown in Figure 6 constitute the PLM global cluster. These are the ‘interoperability’, ‘ontology’, ‘product data management’ and ‘lifecycle assessment’ communities.

These communities are linked together and probably refer to the main challenges that researchers consider they need to address in the PLM domain. Indeed, we identified a community researching PDM-related themes, while PDM has historically been seen as the core of PLM. The data also reveals that interoperability issues are a major challenge in the PLM domain. Moreover, it seems that achieving a full coverage of the product lifecycle can be addressed with ontological approaches, and ‘lifecycle assessment’ is an important aspect of PLM that encompasses the whole product lifecycle.

Interestingly, ‘knowledge management’, while being the top keyword by count, does not form its own community in Figure 5 or in Figure 6. And lastly, we also observe that the results tend to indicate that PLM research is focusing much more on technical issues (such as STEP or OWL) and on their applications’ domains (such as lifecycle costing), while social or cultural aspects are not being discussed in the PLM research community.

7 **Conclusions**

This paper presents a mapping study of the PLM scientific literature published from 2005 to 2015 in peer-reviewed conferences and journals. It establishes a method to analyse a large number of scientific publications in order to characterise a research domain. The approach builds on keywords selected by the authors of the 1,390 retained publications. More precisely, we analysed the co-occurrences of these keywords to understand how different topics and concepts relate to each other. Techniques from social network science were used to analyse the network of 2,947 normalised authors’ keywords connected in pairs via 11,289 edges.

The most critical step in the proposed approach is keyword normalisation. Normalisation can only be managed properly if clear rules are followed. We decided to have a minimal intervention on the data, and thus even keywords that are very similar

remain distinct in our study. As an outcome, we were able to create a profile of PLM in a structured and repeatable process.

Only 745 of the keywords (excluding ‘product lifecycle management’) have a count of two or more. The keyword ‘knowledge management’ is the one with the highest count. The L-shape distribution of the normalised keywords suggests that PLM as a research domain has a coverage that is both wide and thin. Indeed, in the list of 2,947 normalised keywords, 2,201 (74.7%) have a count of just one. Many authors obviously use uncommon keywords.

The visualisation of communities of keywords reveals the structure of the domain better than just the count, degree or similar measures of keywords. In this paper, the PLM communities are revealed by filtering the network of keywords according to the weight of their network edges, rather than by filtering on the keywords’ degrees, as done by Nyffenegger et al. (2016). Hence, the PLM core cluster was obtained first by filtering data so as to show edges with a weight of 4 or more (Figure 5), thus revealing the main research areas of PLM. The PLM global cluster was next revealed by filtering to show edges with a weight of 3 or more (Figure 6), thus adding details about concepts and technologies. The four major communities composing the PLM global cluster are ‘interoperability’, ‘ontology’, ‘product data management’ and ‘lifecycle assessment’. The PLM overall cluster was revealed last by filtering data to show edges with a weight of 2 or more (Figure 7). The PLM overall cluster reveals the ‘intelligent product’ community, which relates to the Industry 4.0 phenomenon. The BIM community is also revealed but remains isolated from the PLM overall cluster.

This mapping study reveals that PLM as a research domain is very widespread and remains hard to define by only a few terms. However, there are a small number of communities that form the core of research activities in PLM; there is a structure in PLM research. There are also a number of satellite topics that may become part of the PLM cluster in the future, such as BIM. In future works, characterising the evolution of these communities over time would be very interesting.

References

- Bhatt, S., Tseng, F.H., Maranzana, N. and Segonds, F. (2015) ‘Scientometric study of product lifecycle management international conferences: a decade overview’, in *IFIP International Conference on Product Lifecycle Management (AICT-467)*, Springer, Cham, pp.672–683.
- Blondel, V.D., Guillaume, J.L., Lambiotte, R. and Lefebvre, E. (2008) ‘Fast unfolding of communities in large networks’, *Journal of Statistical Mechanics: Theory and Experiment*, Vol. 2008, No. 10, p.10008.
- Budgen, D., Brereton, P., Drummond, S. and Williams, N. (2018) ‘Reporting systematic reviews: some lessons from a tertiary study’, *Information and Software Technology*, Vol. 95, pp.62–74, <https://doi.org/10.1016/j.infsof.2017.10.017>.
- Cao, H. and Folan, P. (2012) ‘Product life cycle: the evolution of a paradigm and literature review from 1950–2009’, *Production Planning & Control*, Vol. 23, No. 8, pp.641–662.
- Corallo, A., Latino, M.E., Lazoi, M., Lettera, S., Marra, M. and Verardi, S. (2013) ‘Defining product lifecycle management: a journey across features, definitions, and concepts’, *ISRN Industrial Engineering*, Vol. 2013, <https://doi.org/10.1155/2013/170812>.
- Danjou, C., Rivest, L. and Pellerin, R. (2017) ‘Douze positionnements stratégiques pour l’Industrie 4.0: entre processus, produit et service, de la surveillance à l’autonomie’, *12ème édition du Congrès International de Génie Industriel (CIGI) 2017*, Compiègne, France, 3–5 May.

- Danon, L., Diaz-Guilera, A., Duch, J. and Arenas, A. (2005) 'Comparing community structure identification', *Journal of Statistical Mechanics: Theory and Experiment*, Vol. 2005, No. 9, <https://doi.org/10.1088/1742-5468/2005/09/P09008>.
- David, M. and Rowe, F. (2015) 'Le management des systèmes PLM (product lifecycle management): Un agenda de recherche', *Journal of Decision Systems*, Vol. 24, No. 3, pp.273–297.
- Fink, A. (2013) *Conducting Research Literature Reviews: From the Internet to Paper*, Sage Publications, Los Angeles.
- Grant, M.J. and Booth, A. (2009) 'A typology of reviews: an analysis of 14 review types and associated methodologies', *Health Information & Libraries Journal*, Vol. 26, No. 2, pp.91–108.
- Guimera, R. and Amaral, L.A.N. (2005) 'Functional cartography of complex metabolic networks', *Nature*, Vol. 433, No. 7028, p.895.
- Hart, C. (2018) *Doing a Literature Review: Releasing the Research Imagination*, 2nd ed., Sage Publications Ltd., Los Angeles.
- Hartley, J. and Kostoff, R.N. (2003) 'How useful are 'key words' in scientific journals?', *Journal of Information Science*, Vol. 29, No. 5, pp.433–438.
- Heymann, S. (2006) *Gephi Statistics* [online] <https://github.com/gephi/gephi/wiki/Statistics> (accessed 23 February 2016).
- Holler, M. (2018) *Mapping the Field of Product Lifecycle Management: A Bibliometric Study*, Institute of Information Management.
- Jacomy, M., Venturini, T., Heymann, S. and Bastian, M. (2014) 'ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software', *PLoS ONE*, Vol. 9, No. 6, p.e98679.
- Jalali, S. and Wohlin, C. (2012) 'Systematic literature studies: database searches vs. backward snowballing', in *Proceedings of the ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*, ACM, pp.29–38.
- Kim, C.S., Bai, B.H., Kim, P.B. and Chon, K. (2018) 'Review of reviews: a systematic analysis of review papers in the hospitality and tourism literature', *International Journal of Hospitality Management*, Vol. 70, pp.49–58, <https://doi.org/10.1016/j.ijhm.2017.10.023>.
- Kipp, M.E.I. (2005) 'Complementary or discrete contexts in on-line indexing: a comparison of user, creator and intermediary keywords', *Canadian Journal of Information and Library Science*, Vol. 29, No. 4, pp.419–436.
- Kitchenham, B., Brereton, O.P., Budgen, D., Turner, M., Bailey, J. and Linkman S. (2009) 'Systematic literature reviews in software engineering – a systematic literature review', *Information and Software Technology*, Vol. 51, No. 1, pp.7–15 [online] <https://doi.org/10.1016/j.infsof.2008.09.009>.
- Kitchenham, B.A., Budgen, D. and Brereton, O.P. (2011) 'Using mapping studies as the basis for further research – a participant-observer case study', *Information and Software Technology*, Vol. 53, No. 6, pp.638–651.
- Laumond, J.P. (2006) 'Motion planning for PLM: state of the art and perspectives', *International Journal of Product Lifecycle Management*, Vol. 1, No. 2, pp.129–142.
- Lu, K. and Kipp, M.E.I. (2014) 'Understanding the retrieval effectiveness of collaborative tags and author keywords in different retrieval environments: an experimental study on medical collections', *Journal of the Association for Information Science and Technology*, Vol. 65, No. 3, pp.483–500 [online] <https://doi.org/https://doi.org/10.1002/asi.22985>.
- Mas, F., Arista, R., Oliva, M., Hiebert, B., Gilkerson, I. and Rios, J. (2015) 'A review of PLM impact on US and EU aerospace industry', *Procedia Engineering*, Vol. 132, pp.1053–1060, <https://doi.org/10.1016/j.proeng.2015.12.595>.
- McSweeney, P.J. (2009) 'Gephi network statistics', in *Google Summer of Code 2009 Project Proposal*.

- Ming, X.G., Lu, W.F., Yan, J.Q. and Ma, D.Z. (2005) 'Towards collaborative innovation via product lifecycle management (PLM): status review and technology trend', in *ASME 2005 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, American Society of Mechanical Engineers, pp.817–825.
- Mulrow, C.D. (1994) 'Systematic reviews: rationale for systematic reviews', *BMJ*, Vol. 309, No. 6954, pp.597–599.
- Nappi, V. and Rozenfeld, H. (2013) 'Sustainability performance indicators for product lifecycle management', in *22nd International Congress of Mechanical Engineering (COBEM)*.
- Newman, M.E. (2006) 'Modularity and community structure in networks', *Proceedings of the National Academy of Sciences*, Vol. 103, No. 23, pp.8577–8582.
- Noy, N.F. and Klein, M. (2004) 'Ontology evolution: not the same as schema evolution', *Knowledge and Information Systems*, Vol. 6, No. 4, pp.428–440.
- Nyffenegger, F., Rivest, L. and Braesch, C. (2016) 'Identifying PLM themes, trends and clusters through ten years of scientific publications', in *IFIP International Conference on Product Lifecycle Management*, Springer, Cham, pp.579–591.
- Okoli, C. and Schabram, K. (2010) 'A guide to conducting a systematic literature review of information systems research', *SSRN Electronic Journal* [online] <https://doi.org/10.2139/ssrn.1954824>.
- Petersen, K., Feldt, R., Mujtaba, S. and Mattsson, M. (2008) 'Systematic mapping studies in software engineering', in *EASE*, Vol. 8, pp.68–77.
- Petersen, K., Vakkalanka, S. and Kuzniarz, L. (2015) 'Guidelines for conducting systematic mapping studies in software engineering: an update', *Information and Software Technology*, Vol. 64, pp.1–18, <https://doi.org/10.1016/j.infsof.2015.03.007>.
- Pulkkinen, A., Rissanen, N. and Vainio, V. (2013) 'PLM state of the practice and future challenges in globally networked manufacturing companies', in *Proceedings of the 1st PDM Forum for Finland-Russia Collaboration*, Lappeenranta University of Technology, Lappeenranta, pp.69–75.
- Terzi, S., Bouras, A., Dutta, D., Garetti, M. and Kiritsis, D. (2010) 'Product lifecycle management – from its history to its new role', *International Journal of Product Lifecycle Management*, Vol. 4, No. 4, pp.360–389.
- Varandas Jr., A., Miguel, P.A.C., Carvalho, M.M.D. and Zancul, E.D.S. (2015) 'Product life cycle management and product development: bibliometric analysis and literature classification', *Production*, Vol. 25, No. 3, pp.510–528.
- Webster, J. and Watson, R.T. (2002) 'Analyzing the past to prepare for the future: writing a literature review', *MIS Quarterly*, Vol. 26, No. 2, pp.13–23.
- Wohlin, C. (2014) 'Guidelines for snowballing in systematic literature studies and a replication in software engineering', in *EASE '14 – Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*, ACM, Article No. 38.

Notes

- 1 Web of Science and Science Direct are two of the largest databases for scientific publications.
- 2 About 60 papers published by ASME were not included because no keywords were available.