

Identifying PLM themes, trends and clusters through ten years of scientific publications

Felix Nyffenegger¹, Louis Rivest², Christian Braesch³

¹HSR, Rapperswil, Switzerland
felix.nyffenegger@hsr.ch

²ETS, Montreal, Canada
louis.rivest@etsmtl.ca

³Univ. Savoie Mont-Blanc - SYMME, Annecy, France
christian.braesh@univ-smb.fr

Abstract. PLM encompasses a wide array of expertise, from designing green products to digital factories, with perspectives ranging from an IT standpoint to business strategies, encompassing products, processes and services. Hence, identifying the contours of PLM as a science through the themes, trends and clusters of its scientific literature is very challenging. At the same time, being able to portray PLM will benefit the PLM community, including researchers and practitioners, and should help foresee its future. This work examines PLM research bibliometric trends over the last ten years. We review the scientific literature published in English from 2005 to 2014 in peer-reviewed journals and conferences. Paper keywords are analyzed so as to identify trends and reveal the clusters of related themes based on the occurrences of words and the frequency of associations between them. Amongst the findings we observe that PLM coverage is both very large (2847 keywords being used over a decade) and very thin (2134 of these keywords appear only once in the decade). We also observe that the keyword showing the highest increase is Building Information Modeling (BIM).

Keywords: Bibliometric analysis, PLM, Review, Mapping study, Trends, Overview.

1 Introduction

Product Lifecycle Management (PLM) has gained momentum over roughly the last 15 years, both in industry and in the research community. It can be stated that PLM encompasses a vast array of expertise, from designing green products to continuous manufacturing processes management, and that its perspectives range from Information Technology (IT) and Knowledge representation to Management-oriented works. This variety makes it challenging to define the contour of the very nature of PLM today and where it is headed.

This paper therefore presents a mapping of the PLM scientific literature published from 2005 to 2014 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 last decade and detect underlying trends, if any. This paper also examines some clusters of themes and topics that are connected within PLM by studying the relationships between keywords found in the PLM literature. The first section reviews the literature relevant to our work. The methodology used for paper selection and the analysis technique are presented next, followed by the highlights of our findings. These findings are then analyzed and discussed, and we end with some conclusions and views on future work.

2 Related research

Kitchenham [1] distinguishes *systematic literature reviews* from *mapping studies*. While the former aim at answering specific research questions, mapping studies are suitable to answer general questions, which are more relevant to research trends. Mapping studies aggregate and classify the relevant literature. One important factor for a mapping study is that a stringent search process be used.

Shepperd [2] suggests elaborating systematic reviews in the following steps: problem formulation, locating evidence, appraising evidence quality, evidence synthesis and interpretation and lastly, reporting. Our working methodology implements all of these steps, as described below.

Aside from the scientific literature that may guide our work towards constructing a high quality mapping study, we also searched for general PLM literature reviews or state-of-the-art summaries, and found very few. Bhatt et al. [3] published a PLM review in 2015 identifying some of the most frequent PLM research themes, but it was limited to papers published in PLM International Conference (IC) proceedings.

In 2015, Varandas et al. [4] published an analysis of PLM literature published from 2006 to 2010, with a focus on the concepts of Product Life Cycle Management, New Product Development, Environmental Sustainability and their interfaces (in Portuguese). Also in 2015, Mas et al. [6] 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 fifty years. Just a few years earlier (in 2013), Nappi and Rozenfeld [5] published a PLM literature review focusing on sustainability performance indicators.

Cao & Folan [7] reviewed the evolution of Product Life Cycle literature from 1950 to 2009, showing how it led to PLM. In 2005, Ming et al. [8] also presented a PLM review, which now dates back to over a decade. Some papers have promising titles but their contents do not help us in our work [9], [10].

The excellent paper from Terzi et al. (2010) [11] shares some of our aims and describes the constituent elements of PLM, and classifies them into three fundamentals themes: information and communication technologies, business processes, and methodologies. The findings presented below could clearly be classified along the same lines.

None of the above-mentioned papers exposes global PLM themes, trends and clusters over the years through a systematic PLM literature review based on keywords analysis. Hence, this paper offers a unique, fact-based, perspective on PLM.

3 Research objective

Our objective is to draw a general portrait of PLM scientific literature over the last decade. This general portrait is designed to:

- identify the major themes being studied in the PLM scientific literature, thereby revealing what researchers consider to be part of PLM science;
- identify trends amongst these major themes so as to recognize both the 'rising stars' and those that are in decline over a ten-year period;
- identify in what countries PLM research is being conducted and observe if this research effort is stable, declining or increasing, by country; and
- identify clusters of ideas that are logically connected within the PLM domain.

From there, we should be able to answer, in a next step, questions such as: is CAD (Computer-Aided Design), as a topic, at the heart of PLM? Is BIM (Building Information Modeling) equivalent to PLM but applied exclusively to construction?

The working methodology established to achieve our objectives is described in the next section.

4 Methodology

This study addresses trends in the research fields of PLM and its objectives have a general character. Thus a mapping study was selected as the general methodology to achieve our goal. As noted by Kitchenham [1], particular care has to be taken with a proper selection process to identify the relevant literature and a well-defined and reliable classification system.

We should stress here that even though mapping studies traditionally include a step for the classification of the collected information based on the domain knowledge of experts, the working methodology adopted here avoids any such biased classification. Hence, the general steps suggested by Shepperd [2] for unbiased evidence based studies were followed. The procedures and methods for locating evidence, appraising evidence quality and evidence synthesis and interpretation for this study are explained in the following sections.

For this initial study we decided that the entity of evidence is keywords. We argue that the authors are best qualified to explain their own work based on keywords and, thus, this is an adequate starting point to analyze what topics are being researched and how they are linked to each other. Keywords may be biased by trends or explicit calls for conferences. This effect should be considered when interpreting the results. In upcoming works, the analysis is planned to be extended to abstracts or even full paper texts, rather than keywords only; those efforts would more precisely reveal the publications' contents.

4.1 Paper selection

Paper selection was conducted in three steps. First, a set of selection criteria was established and appropriate databases were searched for relevant papers. The duplicates from different sources were identified and eliminated in the second step. In the third and final selection step, the publication guidelines of the corresponding journals and conferences of all the remaining papers were checked to assure an acceptable scientific standard. These criteria include a neutral (not vendor-influenced) publisher and a peer-review process.

The following selection criteria were used for paper selection:

1. Must be in English;
2. Must have an abstract;
3. Must have keywords¹;
4. Must have a date (at least the year) of publication;
5. Paper must be published between 1.1.2005 and 31.12.2014;
6. Must have PLM or *Product Lifecycle Management* (or any alternative spelling such as Product life-cycle management) in the title, keywords, abstract, or title of the proceedings or journal; and
7. Paper must be peer reviewed in a conference or a journal.

Criteria 6 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.

Papers for this study were selected from four sources:

- International Journal of Product Lifecycle Management (IJPLM);
- Product Lifecycle Management International Conference (PLM IC);
- Web of Science and Science Direct – These databases are two of the largest database for scientific publications.

It can be emphasized that all the papers from the IJPLM and PLM IC satisfy all our criteria, even when a paper does not use PLM in its title or keywords. In addition, keywords added by editors had to be removed for our study, since our hypothesis is based on author’s keywords.

4.2 Appraising and Improving Data Quality

Several steps were taken so as to appraise and improve the quality of our data. As mentioned above, we tried to execute these steps with a minimum of bias. Initially all records were captured into Zotero as a common literature database, where we could identify (and eliminate) duplicates. Next, data exporting was done in order to process the data. Finally, the processed data was merged into the final database with a schema

¹ About 60 papers published by ASME were not included because no keywords were available.

that supports our analysis. The merging and creation of the final database was done automatically by an algorithm developed for that purpose. This algorithm follows the rules identified during the steps of metadata alignment, data enrichment and keyword normalization, as explained below. The automation of this process allows the full process to be repeated in the future, so as to easily include new publications and keep our database up to date.

The most critical step was normalization. The journal and proceedings names, and most importantly, the keywords had to be transferred into a normalized form. This transformation included some interpretation by the authors. However, we conducted this step following a clear set of rules.

- **Metadata alignment** - Metadata from the different sources needed to be aligned, so that equivalent fields of different records have the same meaning, such as, for example, a journal's name or the name of a conference.
- **Metadata enrichment** - Since we aimed to offer a geographic perspective to our data, the publications needed geo location. Hence, each publication was tagged by the country of the first author's institution.
- **Normalization** – A critical step in our analysis was the normalization: Each keyword was translated into a normalized form, so that keywords with the same meaning are unique. We needed to select and apply some rules for this process, including:
 - American English is chosen over Britain English (e.g. 'modelling' is normalized 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.

Particular care was spent on acronyms. The same acronym can have different meanings, and tracing back from the acronym to the correct keyword led at least to searching through a paper's abstract.

4.3 Synthesis and interpretation

An efficient method with which to present aggregated data from different perspectives and at different levels is necessary to allow its interpretation. At this stage we decided on three main perspectives: geographic distribution, a co-occurrence based keyword graph, and the trends of keywords based on their 10 years' of history.

Geographic distribution. As mentioned above, during data quality improvement, we added a country of origin to each publication, according to first author affiliation. Hence, we can provide histograms showing the evolution of the number of publications per year per country over the selected ten-year period.

Building Clusters based on a Co-Occurrence Network. Co-occurrence networks are used to visualize 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 “neighbors” and these are grouped into “neighborhoods” based on their interconnections.

In this study, two keywords become neighbors if they appear in the same publication. A co-occurrence network then helps to visualize the relationships between keywords and highlights clusters of keywords by identifying their neighborhood. As suggested by HuaJiao et al. [12], in our keyword co-occurrence network each keyword is represented by a node. The size of a node correlates to the number of total *occurrences* of a keyword in all publications. Co-occurrence relationships are represented as edges; hence the number of co-occurrences of a given pair of keywords influences the *weight* of this edge. McSweeney [13] and the community around the graph visualization tool Gephi [14] discuss different alternatives to measuring the importance of a node in a graph. For our purpose, the most appropriate measure seems to be the *degree*. The degree corresponds to the number of edges that are attached to a node. Nodes with a high degree play a central role in the network.

To identify clusters, we decided to focus on first-level neighborhoods (all direct co-occurrences of a specific keyword). We used the ForcAtLats2 algorithm [15] to lay out our graph. However, if all nodes and edges are visualized, the graph becomes a so-called “hairball”. To remove complexity from the graph and thereby reveal clusters, controlling the displayed nodes by filtering the degrees and the occurrences of nodes is essential, as shown below.

Analyzing trends. A third perspective on the data is to extract keyword trends. The trend is retrieved by looking at the number of their occurrences per year in a histogram. The linear regression of this histogram gives us an indication about the trend of each keyword. However, the number of occurrences per year per keyword is, in many cases, rather low, so the measure of a trend by linear regression must be looked at critically. In future work, we will extend the study to abstracts or full paper texts, and so the results might be more robust.

5 Some results and findings

5.1 General Findings

Table 1 below summarizes the number of contributions considered in the review according to the criteria described above (peer-review, presence of keywords, etc.), and their sources. We observe that IJPLM is by far the leading PLM journal with 45% of the papers published (146 out of 326), followed by Computers in Industry with 12% (38) of the journal contributions.

Table 1 – Number of scientific contributions and their sources

| Property | Figure |
|---|---|
| Number of publications | 1153 |
| Number of journals involved | 63 |
| Number of publications from journals | 326 |
| Top 3 journals | International Journal of PLM (146) Computers in Industry (38) Computer-Aided-Design (13) |
| Number of conferences involved | 153 |
| Number of publications from conferences | 827 |
| Top 3 conferences | International Conference on PLM (571) CIRP Design Conference (10) International Conference on Digital Enterprise Technology (9) |
| Number of normalized keywords | 2847 |

The PLM International Conference series is the major scientific PLM conference with 69% (571 out of 827) of the conference papers; the second one having published only 1% (10) of these contributions.

5.2 Analysis from Geographical data

Each of the 1153 scientific contributions was associated to the country of the first author's institution. A few observations can be made. First, France is the top contributor with 206 papers over the 2005-2014 period. Bhatt et al. made a similar observation while considering only contributions to the PLM International Conference [3]. Germany is second with 160 contributions. Italy and China are next with 96 contributions each.

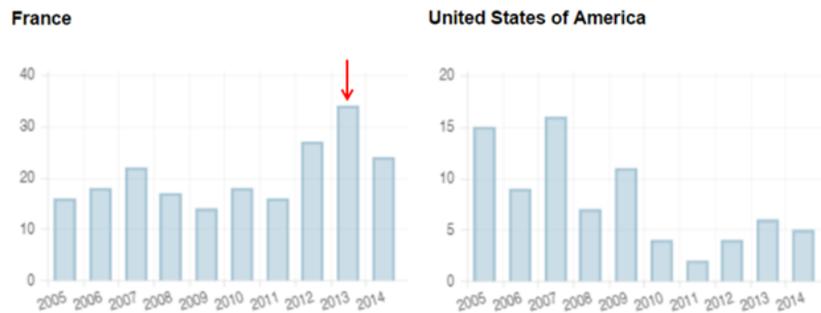


Figure 1. Contributions by France and USA over the 2005-2014 period

Histograms expressing the evolution of contributions by country over time (Figure 1) suggest that scientific production from France is maintained over the ten-year period. The same could be stated for Germany and Italy, even though peaks can be observed. These peaks correspond with the country hosting the PLM IC (France

2013; Germany 2010; Italy 2007). We also note that the number of contributions from the USA seems to have decreased in the second half of the ten-year period.

5.3 PLM Themes as Revealed by Keywords

As indicated in Table 1, the 1153 papers that met our criteria used a total of 2847 normalized keywords. Figure 2 shows the distribution of the normalized keywords degree. The figure shows a distinct L-shape since some keywords are very common, while most of them only occur in a single paper. In fact, 2134 normalized keywords are used only once in the 1153 papers, while 713 were used more than once. Hence, it could be argued that the PLM community is spread both wide and thin.

Table 2 shows the list of the 15 top-ranking keywords, based on their degree, found in the considered literature. We should note that the Keyword *PLM* (normalized 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 IC) where PLM as a keyword may be considered implicit.

Considering that both *ontology* and *knowledge management* refer to **Knowledge Management**, it is reasonable to consider that this is a fundamental topic to PLM (ranks 3 and 4). In a similar manner, *product development*, *product design* and *new product development* all refer to **Product Development**, which is clearly a topic of importance to PLM (ranks 2, 6 and 13). And last and not surprisingly, *interoperability*, *collaboration*, *collaborative design* and *concurrent engineering* all refer to **Collaboration**, which is important to PLM (ranks 8, 9, 11 and 12). It can also be stated that *product data management* and *computer aided design* are two classes of tools that are central to PLM.

These top-ranking keywords describe PLM in a very generic way. Hence, keywords that appear farther down on the list are ‘less obvious’ and may describe second-level ideas. For example, *product lifecycle management system* (rank 14) is a subset of *product lifecycle management*. Thus, filtering out the top level keywords helps unveil the less obvious topics. This filtering is illustrated later where we show how the *computer aided design* cluster is revealed.

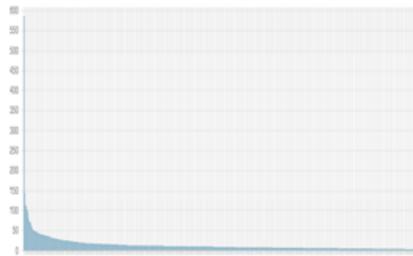


Figure 2. Distribution of normalized keywords degree

Table 2 – Top-ranking normalized keywords and their degree

| Rank | Name | Degree |
|------|-------------------------------------|--------|
| 1 | product lifecycle management | 585 |
| 2 | product development | 144 |
| 3 | ontology | 114 |
| 4 | knowledge management | 111 |
| 5 | product data management | 109 |
| 6 | product design | 101 |
| 7 | computer aided design | 101 |
| 8 | interoperability | 93 |
| 9 | collaboration | 80 |
| 10 | product lifecycle | 74 |
| 11 | collaborative design | 72 |
| 12 | concurrent engineering | 70 |
| 13 | new product development | 65 |
| 14 | product lifecycle management system | 58 |
| 15 | lifecycle assessment | 58 |

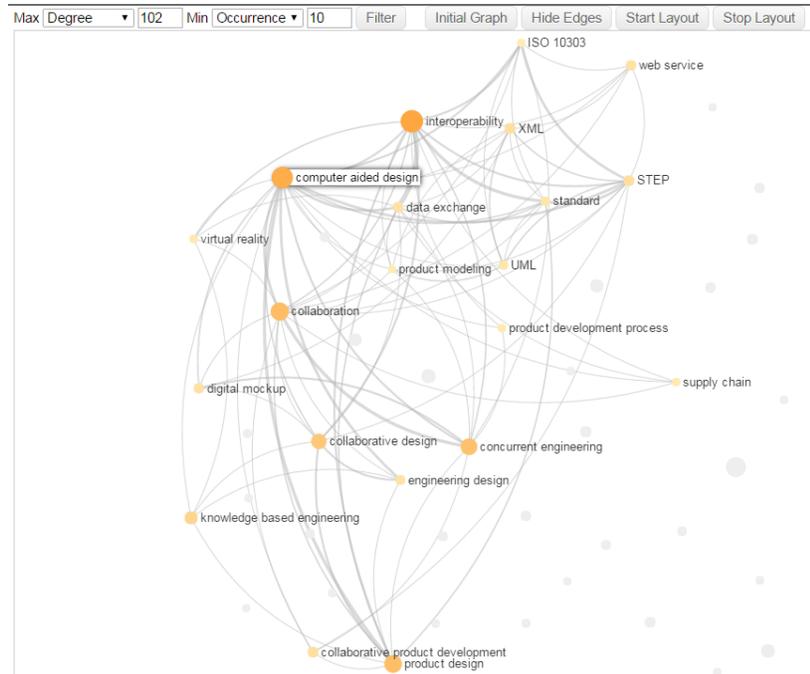


Figure 5. Computer aided design cluster of keywords

Further filtering helps to reveal clusters. For example, Figure 5 shows the keywords with a degree less than 102 (which serves to hide the five top keywords from Table 2) and an occurrence superior to 10. This tool reveals the clusters of keywords associated with a specific one. The first-level neighbors of *computer aided design* are displayed, hence revealing the CAD cluster. Interestingly, but not surprisingly, this cluster includes *STEP*, *digital mockup*, *data exchange*, *collaboration* and *product design*, amongst others.

Figure 6 shows the cluster around *building information modeling*, previously identified as a rising star. This cluster reveals that *building information management* has a strong connection with *building information modeling*, and that *digital mockup* is connected with *maintenance* (but not with design), suggesting that the considered scientific literature sees *digital mockup* as a tool for *building maintenance*, but that using digital mockups for building design is not explored. Of course, the reader should remember that the BIM literature surveyed here is limited to a small number of papers that were PLM-tagged so as to be included in our work. Hence, this cluster tells a story about BIM/PLM research, one that would probably differ if based on ‘pure’ BIM literature.



Figure 6. *Building information modeling* cluster of keywords

6 Conclusion and some future work

This paper presents a method to analyze a large corpus of 10 years of scientific publications in order to characterize PLM. The 1153 publications included in this corpus were filtered according to a list of criteria, such as peer-review. The proposed approach builds on keywords selected by the authors of these publications. These keywords were normalized to limit variations due to American or British English, and so on. That process yielded a total of 2847 normalized keywords, 2134 of them used only once. It can be concluded from these numbers, and from the L-shape distribution of the normalized keywords, that PLM as a research domain has a coverage that is both wide and thin. It would be interesting to conduct similar analysis on other research domains so as to compare the distribution of keywords.

Characterizing PLM through normalized keywords also allows to observe the evolution of keywords over a ten-year period, and to identify the keywords that are gaining or declining in popularity. The data clearly shows, for example, that ‘building information modeling’ is now being considered as a part of PLM by its community of researchers.

The presented work also used the co-occurrences of keywords in a given paper so as to capture relations between keywords. If multiple papers use the same pair of normalized keywords, the intensity of this relation is reflected in this association. We are therefore able to identify keywords that belong to the same cluster as well as the intensity of the pairs within clusters. In future work we will characterize clusters related to any significant keyword and even compare clusters. We can even envision characterizing the evolution of clusters over time.

One challenge of the selected approach is that the number of occurrences of keywords is low and hence compromises the statistics and trends. However, this

limitation could be overcome by conducting a similar analysis of abstracts and even full-text analyses. Amongst the other steps of this ongoing project, publications from 2015 will be added, and the graphic analysis tools will eventually be made available to colleagues through a dedicated website.

References

1. Kitchenham, B.A., Budgen, D., Brereton, P.: The Value of Mapping Studies: A Participant-observer Case Study. In: Proceedings of the 14th International Conference on Evaluation and Assessment in Software Engineering, British Computer Society (2010)
2. Shepperd, M.: Combining Evidence and Meta-analysis in Software Engineering, in: Software Engineering, Springer Berlin Heidelberg, pp. 46–70. (2013)
3. Bhatt, S., Tseng, F.H., Maranzana, N., Segonds, F.: Scientometric Study of Product Lifecycle Management International Conferences: A Decade Overview, (2015)
4. Varandas Junior, A., Miguel, P.A.C., Carvalho, M.M., Zancul, E.S.: Product life cycle management and product development: bibliometric analysis and literature classification, *Production*, vol.25, no.3, São Paulo, pp. 510-528. (2015)
5. Nappi, V., Rozenfeld, H.: Sustainability performance indicators for product lifecycle management. 22nd International Congress of Mechanical Engineering (COBEM) (2013)
6. Mas, F., Arista, R., Oliva, M., Hiebert, B., Gilkerson, I., Rios, J.: A Review of PLM Impact on US and EU Aerospace Industry, *Procedia Engineering*, vol. 132, pp. 1053-1060.(2015)
7. Cao, H., Folan P.: Product life cycle: the evolution of a paradigm and literature review from 1950–2009, *Production Planning & Control: The Management of Operations* vol. 23, Issue 8, pp. 641-662 (2012)
8. Ming, X. G., Lu, W. F., Yan, J. Q., Ma, D. Z.: Towards Collaborative Innovation Via Product Lifecycle Management (PLM): Status Review and Technology Trend, ASME 2005 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (2005)
9. Laumond, J.P.: Motion planning for PLM: state of the art and perspectives, *International Journal of Product Lifecycle Management (IJPLM)*, vol. 1, no. 2. (2006)
10. Antti, P., Rissanen, N., and Vainio, V.: PLM state of the practice and future challenges in globally networked manufacturing companies. The 1st PDM forum for Finland-Russia collaboration, Lappeenranta. (2013)
11. Terzi, S., Bouras, A., Dutta, D., Garetti, M., Kiritsis, D.: Product lifecycle management – from its history to its new role, *IJPLM*, vol.4, no 4, pp 360-389 (2010)
12. Li, H. An, H., Wang, Y., Huang J., Gao, X.: Evolutionary features of academic articles co-keyword network and keywords co-occurrence network: Based on two-mode affiliation network, *Physica A: Statistical Mechanics and its Applications*, vol. 450. (2016)
13. McSweeney, P. J.: Gephi Network Statistics, in: Google Summer of Code 2009 Project Proposal (2009)
14. Heymann, S.: Gephi Statistics, <https://github.com/gephi/gephi/wiki/Statistics> (visited on 23/02/2016)
15. Jacomy M., Venturini T., Heymann S., Bastian, M.: ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. *PLoS ONE* 9(6): e98679. doi:10.1371/journal.pone.0098679 (2014)