SURVEY PAPER

CrossMark

Big data in manufacturing: a systematic mapping study

Peter O'Donovan[*], Kevin Leahy, Ken Bruton and Dominic T. J. O'Sullivan

* Correspondence:
peter_odonovan@umail.ucc.ie
IERG, University College Cork, Cork,
Ireland

Abstract

The manufacturing industry is currently in the midst of a data-driven revolution, which promises to transform traditional manufacturing facilities in to highly optimised smart manufacturing facilities. These smart facilities are focused on creating manufacturing intelligence from real-time data to support accurate and timely decision-making that can have a positive impact across the entire organisation. To realise these efficiencies emerging technologies such as Internet of Things (IoT) and Cyber Physical Systems (CPS) will be embedded in physical processes to measure and monitor real-time data from across the factory, which will ultimately give rise to unprecedented levels of data production. Therefore, manufacturing facilities must be able to manage the demands of exponential increase in data production, as well as possessing the analytical techniques needed to extract meaning from these large datasets. More specifically, organisations must be able to work with big data technologies to meet the demands of smart manufacturing. However, as big data is a relatively new phenomenon and potential applications to manufacturing activities are wide-reaching and diverse, there has been an obvious lack of secondary research undertaken in the area. Without secondary research, it is difficult for researchers to identify gaps in the field, as well as aligning their work with other researchers to develop strong research themes. In this study, we use the formal research methodology of systematic mapping to provide a breadth-first review of big data technologies in manufacturing.

Keywords: Big data; Manufacturing; Smart manufacturing; Industry 4.0; Big data analytics; Engineering informatics; Machine learning; Big data systems; Distributed computing; Cyber physical systems; Internet of things, IoT

Introduction

Modern manufacturing facilities are data-rich environments that support the transmission, sharing and analysis of information across pervasive networks to produce manufacturing intelligence [1–3]. The potential benefits of manufacturing intelligence include improvements in operational efficiency, process innovation, and environmental impact, to name a few [4, 5]. However, similar to other industries and domains, the current information systems that support business and manufacturing intelligence are being tasked with the responsibility of storing increasingly large data sets (i.e. Big Data), as well as supporting the real-time processing of this 'Big Data' using advanced analytics [5–10]. The predicted exponential growth in data production will be a result of an increase in the number of instruments that record measurements from physical environments

O'Donovan *et al. Journal of Big Data* (2015) 2:20

Page 2 of 22

and processes, as well as an increase in the frequency at which these devices record and persists measurements. The technologies that transmit this raw data will include legacy automation and sensor networks, in addition to new and emerging paradigms, such as the Internet of Things (IoT) and Cyber Physical Systems (CPS) [1, 11, 12]. The low-level granular data captured by these technologies can be consumed by analytics and modelling applications to enable manufacturers to develop a better understanding of their activities and processes to derive insights that can improve existing operations.

The focus on big data technologies in manufacturing environments is a relatively new interdisciplinary research area which incorporates automation, engineering, information technology and data analytics, to name a few. At this point in time, it is important to understand the current state of the research pertaining to big data technologies in manufacturing, and to identify areas where future research efforts should be focused to support the next-generation infrastructure and technologies for manufacturing. Therefore, this study aims to classify current research efforts, derive prominent research themes, and identify gaps in the current literature.

This study employs the well-known and formal secondary research method of systematic mapping to capture the broad and diverse research strands currently related to big data technologies in manufacturing [13]. The contribution of this study is a comprehensive report on the current state of research pertaining to big data technologies in manufacturing, including (a) the type of research being undertaken, (b) the areas in manufacturing where big data research is focused, and (c) the outputs from these big data research efforts. The research methodology employed in this study is guided by the systematic mapping process described by Peterson et al. [13].

The remainder of this paper is described as follows. In section 2, the research methodology and process used in the study (i.e. systematic mapping) are described. In section 3, the results of the study are presented. In section 4, the study results are discussed in detail. In section 5, the threats relating to the validity of the results are considered. Finally, in section 6 the conclusions from the research are presented and future areas of research are identified.

Research methodology

This study employed systematic mapping to capture the current state of the research relating to big data technologies in manufacturing. Compared with other secondary research methods, such as traditional literature reviews, a mapping study provides an approach that facilitates an investigation of great breadth, while sacrificing depth [13]. In the context of this study, a systematic mapping method was deemed appropriate as it provided a formal and well-structured approach to synthesising material. This structure also served to provide a foundation for reducing bias and harmonising literature review efforts across the research team. Furthermore, the breadth-first perspective that can be derived by systematic mapping was especially useful for reporting on a new and pervasive area of research (i.e. big data in manufacturing) that currently lacks prominent and consistent theories. Indeed, it is the lack of strong research themes that makes a depth first literature review of the area a challenging undertaking.

The mapping process

Figure 1 provides a visual workflow of the systematic mapping process that was used in this study. The main process steps are shown at the top, with each steps outcome shown at the bottom. As research efforts progress through the process, the outcome from each step forms the input for the next step.

At the beginning of the study, the initial research questions were agreed to provide a general scope for the study. Based on this scope, primary search terms and phrases were identified and used to find research papers listed in several digital databases. After the results of these searches were recorded, each paper was manually screened using a set of inclusion and exclusion criteria that attempts to identify papers that are aligned with the theme and scope of the study. Those papers that were deemed relevant to the study were further analysed to determine prominent keywords and phrases that could be used to classify the research being conducted in the area. Finally, the classified papers were aggregated, visualised and mapped in a manner that would enable us to answer the research questions posed in this study.

Research questions

The purpose of this research is to classify current research, and identify pertinent themes and trends in the literature, which relate directly to big data technologies in manufacturing. The main research question is intentionally broad as there are currently no substantial reviews or surveys that address the question, and it is the intention of this study to establish a map that that will convey the diverse themes associated with the area. Therefore, the guiding research question of the study is:

"How are big data technologies being used in manufacturing?"

To answer the main research question, five ancillary research questions that relate to various aspects of big data in manufacturing were identified. Decomposing and characterising the main research question in to smaller and more specific questions enables the topic to be considered from multiple perspectives, while also providing the results needed to answer the main research question. The additional research questions are described below.

RQ1: What is the publication fora relating to big data in manufacturing?

Rationale: The intention of this question is to illustrate the interest in the research area over time, as well as identifying the primary sources of literature in the field. This study assumes that the publication rate is indicative of research interest in the area, while the most prominent sources of research in the field are those journals and conferences that have the highest publication frequency of relevant literature.
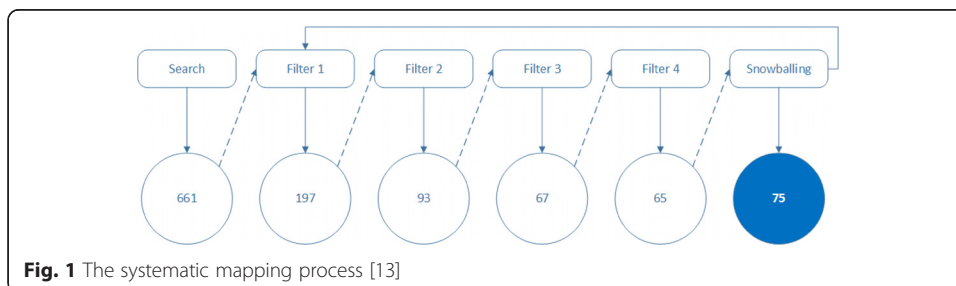

Fig. 1 The systematic mapping process [13]

O'Donovan *et al. Journal of Big Data* (2015) 2:20

Page 4 of 22

RQ2: What type of research is being undertaken in the area of big data in manufacturing?

Rationale: The intention of this question is to highlight the type of formal research being undertaken in the area, ranging from philosophical perspectives, to real-world evaluations. By answering this question the study aims to understand the maturity level of the research area, with the assumption that research efforts that do not exhibit rigorous validation and evaluation may be indicative of a field that is still maturing and focused on developing methodologies to support future research efforts.

RQ3: What type of contributions are being made to the area of big data in manufacturing?

Rationale: The intention of this question is to understand the type of contributions and outputs from research efforts in the field. These outputs may vary greatly and range from information system architectures, to analytical tools and methods for process optimisation. By answering this question the study aims to further assess the maturity level of the field, with the assumption that early research efforts may focus on guidelines and methodologies, and more mature research areas may focus on implementing, evaluating and validating these methods. Furthermore, identifying trends and patterns in the research outputs in the field will also provide an understanding to the approaches used to solve specific challenges in the area.

RQ4: What type of analytics are being used in the area of big data in manufacturing?

Rationale: The intention of this question is to identify the prominence of big data analytics in the research relating to big data technologies in manufacturing, as well as classifying the specific type of analytics being used. In recent years, the term analytics has become synonymous with big data technologies. By answering this question the study aims to better understand the relationship between analytics and big data in the context of manufacturing. Furthermore, the classification of the different types of big data analytics used in research can provide an understanding of the types of problems being addressed.

RQ5: What areas of manufacturing are big data technologies being applied?

Rationale: The intention of this question is to highlight the different areas in manufacturing facilities where researchers are employing big data. By answering this question the study aims to highlight specific research themes, as well as identifying the areas of manufacturing operations that are striving to meet the challenges of large-scale data production and processing.

Primary search

Due to the focus of this study, the search terms 'big data' and 'manufacturing' were considered to be the most obvious primary search terms. Table 1 shows candidate search terms

Table 1 Main and candidate search terms for big data in manufacturing

Main terms	Candidate terms
Manufacturing	Smart Manufacturing, Advanced Manufacturing, Industry 4.0, Cyber Physical Systems, Supply Chain, Factories, Factory, Production, and Process.
Big Data	Large-scale Data, Cloud Computing, Machine Learning, Big Data Analytics, Data Virtualization, and Master Data Management.

Table 2 Primary search string used for study

(Manufactur* OR Factor* OR Industry 4.0)
AND
(Big Data)

identified through the exploration of paper abstracts that were returned by the search query using both 'big data' and 'manufacturing'.

After evaluating different combinations of the candidate search terms, the primary search string was amended (Table 2) to include 'Factory' and 'Industry 4.0'. Our evaluation of different search strings showed that the results that appeared to be most relevant to the study consistently contained the term 'big data', as well as one or more of the terms 'manufacturing', 'factory' or 'factories', in their title, abstract or keywords section. The wild card symbol (*) was used to capture changes in pluralisation and context for relevant populations. The rationale behind the primary string selection was to keep the search broad to capture as many research themes and trends as possible, while also omitting research papers that were not directly aligned with the scope and theme of the study.

The chosen primary search string was used as the search criteria in seven well-known digital databases. These databases were chosen collectively by all researchers involved in the study, and were deemed a relevant source of technology, engineering and industrial research. As each database's search facility is different, the primary search string had to be transformed to the native syntax of each database. In addition, search results were limited to journal and conference publications based on the assumption that these publications are more likely to be peer-reviewed than other sources, such as white papers and book chapters. Table 3 outlines the digital databases that were searched during the study, along with the number of publications returned using the primary search string. The only database that did not have the facility to restrict the search criteria to both title and abstract was Google Scholar. Therefore, two numbers have been included for Google Scholar – 9540 relates to the number of documents returned for the primary search string, while 14 relates to the number of documents returned for the primary search string when limited to the documents title. Given the results from the other digital databases, the lower number would seem to be a more realistic result.

Screening of research

The primary search results were filtered using a set of inclusion and exclusion criteria to identify the most relevant research for the study. Figure 2 provides a breakdown of the research included at each stage in the screening process. Firstly, the primary search

Table 3 Search results from digital repositories

Digital database	Number of publications
ACM Digital Library	283
Engineering Village	128
Google Scholar	14 (9540)
IEEE Xplore	86
Science Direct	14
Scopus	96
Web of Science	40

O'Donovan *et al. Journal of Big Data* (2015) 2:20

Page 6 of 22

Fig. 2 A breakdown of the screening process
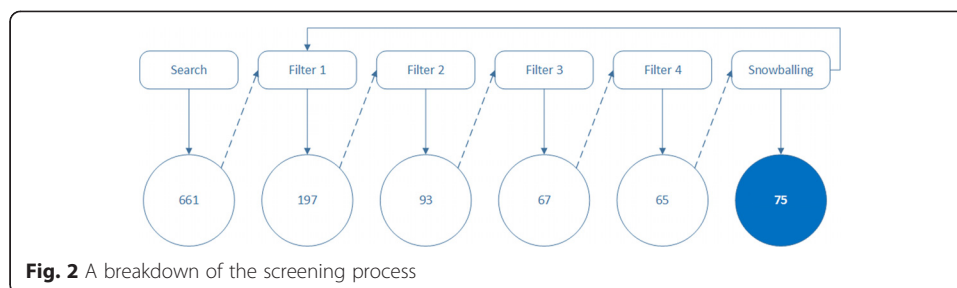
string was used in each of the digital repositories shown in Table 3, which yielded 661 publications. Secondly, these papers were processed using four filters with the intention of omitting publications that were not highly relevant to the study, which resulted in 65 publications remaining. Finally, the snowballing method [13, 14] was used to extract the references from each of these publications, with each reference being screened to ascertain if the research should be added to the study. This resulted in 75 publications being used as the basis for the data in the study.

Inclusion criteria

To be considered for inclusion in the study, the research being evaluated had to originate from an academic source, such as a journal or conference, and clearly show its contribution was focused on big data in manufacturing, which was primarily determined by the presence of the primary search terms. Publications that met this criteria were then processed using exclusion criteria (i.e. filters), with the intention of highlighting the most relevant research in the area of big data in manufacturing.

Exclusion criteria

For those publications that passed the inclusion criteria, four filters were applied to reduce the publications to only those that were deemed to be directly aligned with the focus of the study. These filters are described as follows;

- Filter 1: remove publications that do not contain 'manufacturing', 'factory', 'factories' or 'Industry 4.0' in the title, abstract or meta-data section of the document.
- Filter 2: remove publications that do not contain 'big data' in the title, abstract or meta-data sections of the document.
- Filter 3: remove papers that only refer to 'manufacturing', 'factory', 'factories', 'Industry 4.0', or 'big data' as a fleeting point of reference. For example, many big data related papers cite the potential application of big data to manufacturing, without exclusively investigating the area.
- Filter 4: review the introduction and discussion sections of each publication, and remove those that do not focus on, and contribute to, the area of big data in manufacturing.

Classification of research

All of the publications in the study were classified using four dimensions. These dimensions were chosen to provide different perspectives on the current state of research in

O'Donovan *et al. Journal of Big Data* (2015) 2:20

Page 7 of 22

the area, while also building a data set that could be used to answer each of the research questions highlighted in the study.

Type of research

To classify the type of research approaches being undertaken an existing classification scheme was chosen. This scheme was defined by Wieringa et al. [15] for classifying requirements engineering research. Table 4 provides a summary of each type of research.

Area in manufacturing

To classify the different areas of manufacturing an existing classification scheme was chosen. This scheme was defined by Meziane et al. [16] when identifying different areas of manufacturing where Artificial Intelligence (AI) was being applied. Table 5 provides a summary of each area of manufacturing.

Type of contribution

To classify the type of contribution made by each publication the qualitative research method known as keywording [13] was chosen. The abstract of each publication was synthesised to highlight the main keywords that described the contribution of the research. These words were aggregated and synonymous terms were amalgamated to simplify results and promote better visibility of trends. In terms of cross-overs between keywords, there were significant similarities between research relating to systems and platforms, which resulted in both being labelled as platforms. Similarly, research labelled as guidelines, roadmaps and requirements were unified as theory. Table 6 provides a summary of each type of research contribution. Furthermore, each publication used in the study is classified in Appendix.

Type of analytics

To classify the type of analytics an existing classification scheme was chosen. This scheme was defined by Delen et. al [17] as a high-level taxonomy for business analytics. Table 7 provides a summary of each type of analytics.

Results

RQ1 – What is the publication fora relating to big data in manufacturing?

Figure 3 illustrates the year-on-year growth in publications relating to big data in manufacturing. The first publication identified in this study is from 2009, with no further activity until 2012. From 2012 to 2014 the publications in the field increased almost ten-fold,

Table 4 Types of research [15]

Classification	Description
Validation	Research that investigates novel and unique techniques but have not yet been implemented in real-world environments.
Evaluation	Research that includes a significant implementation of a given technique along with a complete evaluation.
Solution	Research that includes an illustration or example of a solution to a particular problem.
Philosophical	Research that provides a conceptual way of looking at a particular problem or field.
Opinion	Research that expresses a personal opinion about whether a particular technique is good or bad, without focusing on related work or standard research methods.
Experience	Research that is written from the personal experience of the researcher, and describes how something was done.

Table 5 Areas in manufacturing [16]

Classification	Description
Design	Research focusing on the design of product for manufacturing activities.
Process and Planning	Research focusing on all aspects of process and planning, with a core emphasis on the reduction of waste and the increase of output yielded.
Quality Management	Research focusing on quality management in manufacturing environments.
Maintenance and Diagnosis	Research focusing on the health of machinery in manufacturing operations, ranging from predictive maintenance, to real-time diagnostics.
Scheduling	Research focusing on the scheduling, management and optimisation of activities and processes in manufacturing environments.
Control	Research focusing on the control, management and optimisation of operations and processes in manufacturing environments.
Environment, Health and Safety	Research focusing on the factors relating to the environment, energy, as well as health and safety.
Virtual Manufacturing	Research focusing on the realisation of virtual factories and processes.

with a 180 % increase in publications between 2012 and 2013, and a 242.9 % increase between 2013 and 2014. It should be noted that the decrease in publications for 2015 is incomplete, as the data from this study only includes research published in January and February of 2015.

Figure 4 provides a breakdown of publications by journal and conference. Excluding the partial data for 2015, conference publications were greater than that of journal publications for each year that was illustrated. Indeed, as interest in the area began to increase from 2012 to 2014, the proportion of conference to journal publications rose from 60 % in 2012 to 75 % in 2014. As this area of research is relatively new, there is an inherent limitation on the amount of historical data available to identify longstanding and strong correlations, however, based on the data available there is a strong linear correlation between conference and journal publication growth.

Figure 5 highlights the distribution of publications by journal and year. There are four journals that are responsible for publishing 45.84 % of the research in the area. The International Journal of Production Economics is the top source of research in the area with 16.67 % of publications, with the Journal of Business Logistics publishing 12.5 %, while Manufacturing Letters and Studies in Computational Intelligence have published 8.34 and 8.33 % respectively. In addition to the fact that the Journal of Production Economics has the

Table 6 Types of contribution

Classification	Description
Architecture	Research that provides a theoretical view of how various components in a solution will sit together and interact.
Framework	Research that describes the encapsulation of multiple software libraries that solve a particular problem, while also being extensible.
Theory	Research that develops high-level guidelines and roadmaps for a particular problem.
Methodology	Research that presents low-level approaches to solving a particular problem.
Model	Research that produces mathematical models for solving particular problems.
Platform	Research that provides a system with hardware and software components, which enables applications to execute.
Process	Research that presents low-level processes to solving a particular problem.
Tool	Research that develops well-defined software utilities that address a subset of a bigger problem.

Table 7 Types of analytics [17]

Classification	Description
Descriptive	Research that is focused on describing the structure, relationships and meaning of data.
Predictive	Research that is focused on predicting an outcome using the available data.
Prescriptive	Research that is focused on prescribing actions using the available data.

most publications relating to big data in manufacturing, it may also be deemed as the most current, with all of its research published in 2015.

Figure 6 provides a visual summary of publications by conferences and year. The top three sources identified in this study constitute 30.77 % of total conference publications. The IEEE Conference on Big Data is the top source of research with 11.54 % of publications, while the Winter Simulation Conference is the third most prominent source with 7.69 %. The second most prominent source of research is Procedia CRIP with 11.54 % of all conference publications – but it should be noted that Procedia CRIP is an aggregated source of publications from all CRIP conferences.

RQ2 – What type of research is being undertaken in the area of big data in manufacturing?

Figure 7 provides a breakdown of the type of research being conducted in the area of big data in manufacturing. At 52.63 %, the majority of research conducted is philosophical, which is deemed as research that does not aim to demonstrate the concept(s) presented in the paper. The next most common type of research is solution-based research, which solves a particular problem and presents a small example of the solution in practice. Less prominent types of research include evaluation and validation, which each comprise 11.84 and 5.26 % of publications respectively. Evaluation-based research can be considered an advanced type of solution, which explores the full consequences, benefits and drawbacks of the proposed method or approach.

Figure 8 illustrates the popularity of research type by year. Rather predictably, due to the popularity of evaluation and solution research highlighted in Fig. 7, the first and second types of research papers published between 2012 and 2014 are evaluation and solution respectively. Early research efforts in 2012 possessed a strong
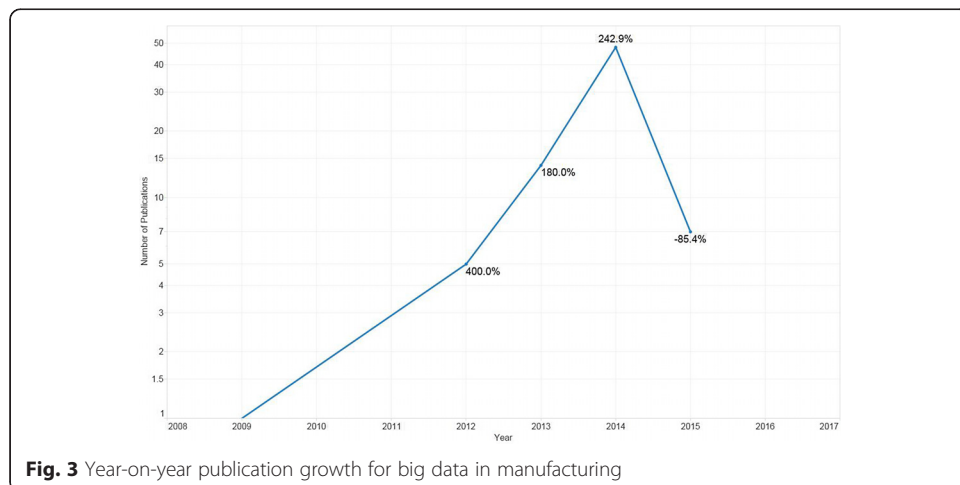

Fig. 3 Year-on-year publication growth for big data in manufacturing

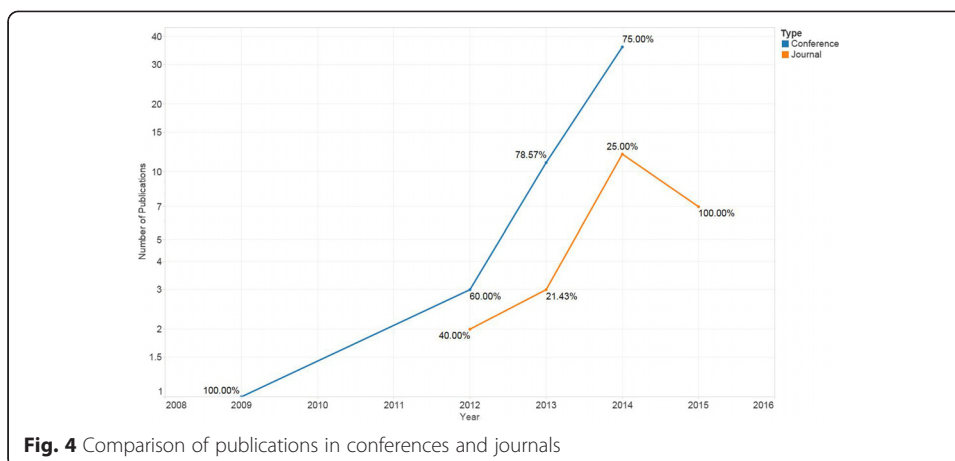O'Donovan *et al. Journal of Big Data* (2015) 2:20

Page 10 of 22

Fig. 4 Comparison of publications in conferences and journals

focus on philosophical research, comprising 60 % of the papers published. As interest in the area grew between 2013 and 2014, the percentage of papers focused on developing philosophies decreased overall, declining to 42.86 % in 2013, and then rising again to 55.1 % in 2014. The trend in publications relating to solution-based research are closely aligned with that of philosophical-based research, comprising 35.71 and 34.69 % of the overall publications in 2013 and 2014. There also appears to be a visible lag in evaluation-based research, which only comprised 4.08 % of publications in 2014, but the partial data for 2015 shows that number of publications in the first quarter of 2015 is twice that of 2014.

RQ3 – What type of contributions are being made to the area of big data in manufacturing?

Figure 9 illustrates the types of research outputs and contributions, as well as decomposing these contributions by conference and journal publications. The top three types of contributions constitute 61.33 % of all publications. At 26.67 %, the most prominent type of contribution is theory. Indeed, this data aligns well with the previous results from Fig. 7, which highlighted philosophical-based research as the most common type
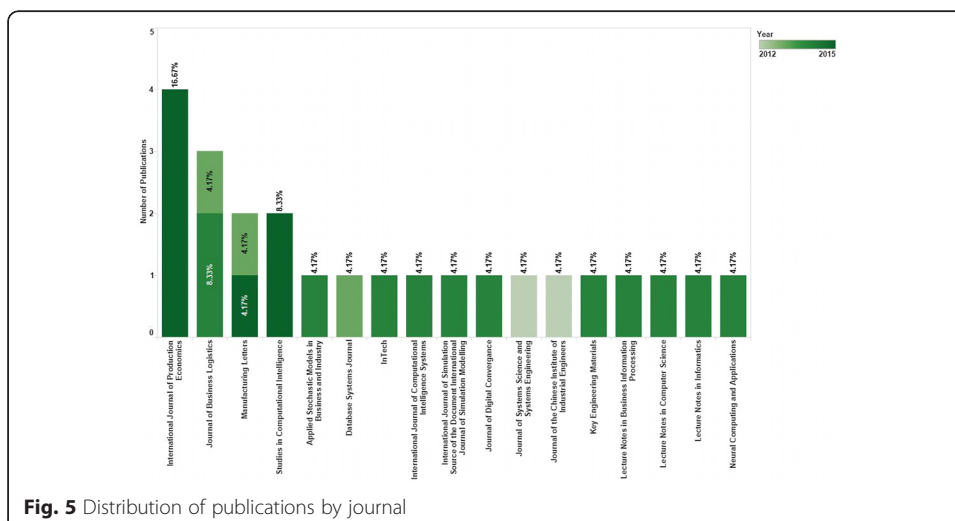

Fig. 5 Distribution of publications by journal

O'Donovan *et al. Journal of Big Data* (2015) 2:20
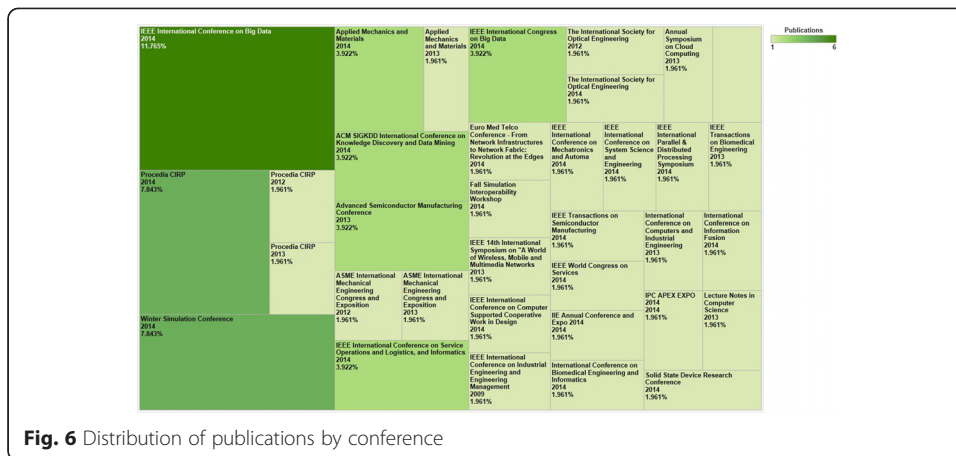
Page 11 of 22

Fig. 6 Distribution of publications by conference

of research being conducted. The next most significant contributions are frameworks and platforms, with each type of contribution associated with 17.33 % of all publications. However, one oddity that is evident in the data is the disproportionate amount of conference to journal papers associated with platforms. This trend is not one that is present in any of the significant contribution types, where expected conference to journal ratios predictably fall in favour of conferences, but with a healthy distribution between both types of publication nonetheless.

Figure 10 illustrates the popularity of research contributions by year. Similar to the overall trend in research contributions shown in Fig. 9, the year-by-year trend in research contributions reinforces the view that theory is the most prominent contribution in the area, with theory being the most prominent contribution from 2012 to 2014. A notable trend is the increased diversity and distribution of research contributions year-on-year, growing from two types of contribution in 2012, to a total of eight in 2014. This is presumably a by-product of increased publication rates, but the distribution provides a balanced view of the research current contributions in the field.

RQ4 – What type of analytics are being used in the area of big data in manufacturing?

Figure 11 shows the percentage of research included in this study that possessed a reference to analytics and big data in manufacturing. The results are relatively even, with 47.69 % of the papers employing some form of analytics, and 52.31 % focusing solely on big data technology without addressing analytics directly.

Figure 12 illustrates the popularity of the main types of analytics on a year-by-year basis. The trend from 2009 to 2015 is remarkably clear. The majority of analytics focus on
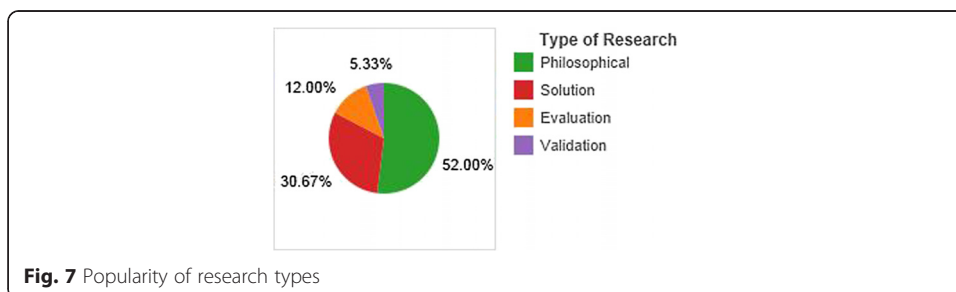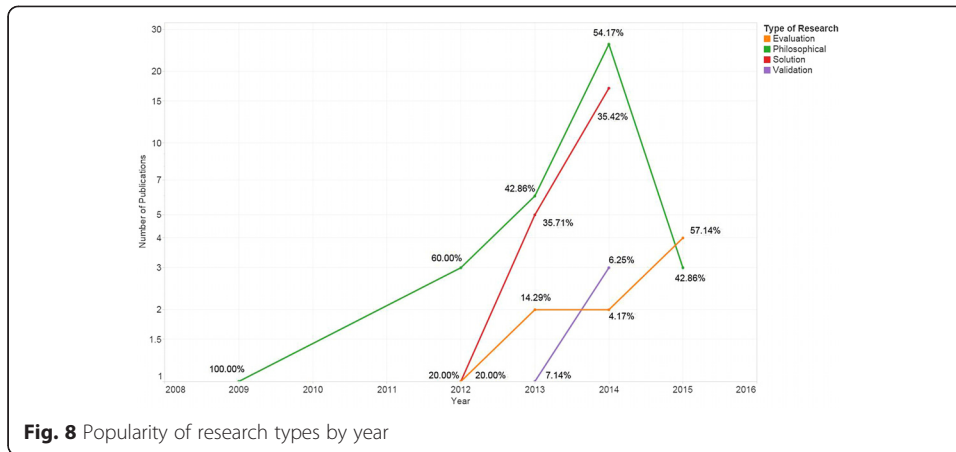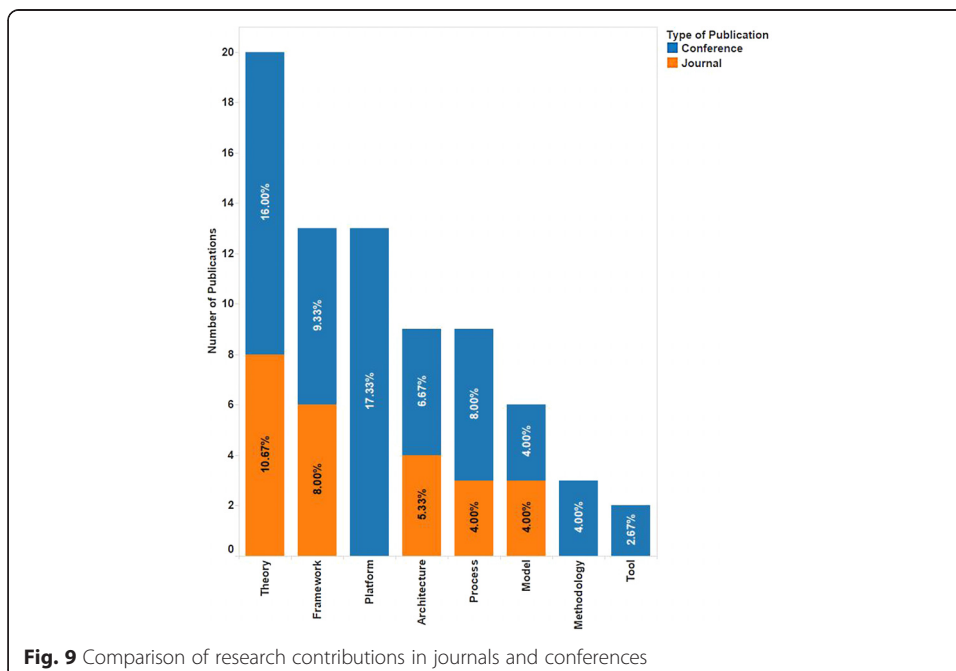

Fig. 7 Popularity of research types

O'Donovan *et al. Journal of Big Data* (2015) 2:20

Page 12 of 22

Fig. 8 Popularity of research types by year

predictive analytics, with a minority focused on descriptive analytics. At the peak publication level in 2014, some form of predictive analytics was evident in 71.43 % of publications, compared to descriptive analytics at 25 %. In addition to these upward trends, a notable stagnant trend is the lack of focus on prescriptive analytics, which accounted for 3.57 % of publications in 2014. Indeed, only a single paper was published in each year between 2012 and 2014, which focused on prescriptive analytics.

RQ5 – What areas of manufacturing are big data technologies being applied?

Figure 13 shows areas in manufacturing where research efforts are using big data technologies, as well as identifying significant contributions within each area. As is evident from the findings, process and planning is the most prominent area of manufacturing for research pertaining to big data technologies. Within process and planning, the most prominent contributions are architectures, platforms, frameworks and theories, comprising over 60 % of
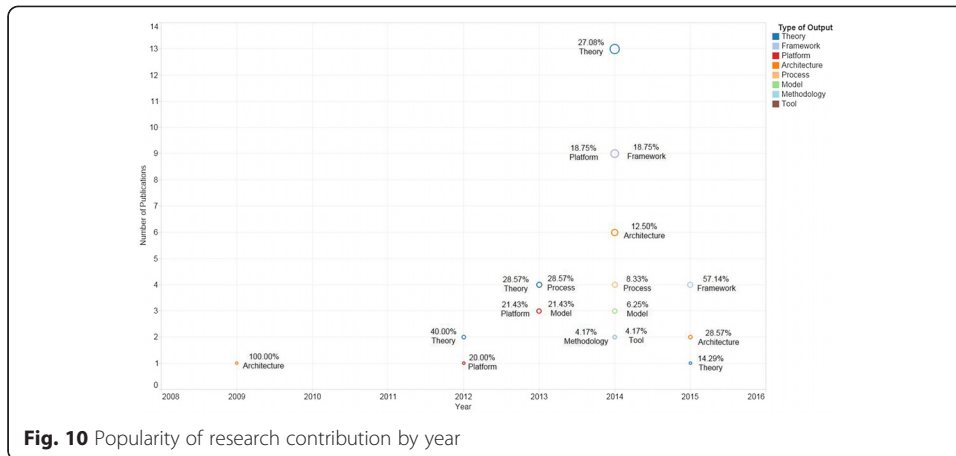

Fig. 9 Comparison of research contributions in journals and conferences

O'Donovan *et al. Journal of Big Data* (2015) 2:20

Page 13 of 22

Fig. 10 Popularity of research contribution by year

publications in that particular area. Following process and planning is enterprise, which is a high-level label that represents research that spans multiple areas in manufacturing, or research that focuses on a single area but cannot be clearly attributed to any single class. Contributions associated with the enterprise share a similar distribution to process and planning, where frameworks, theories and processes are all significant contributions from research in the area. Given that enterprise is an aggregate of sorts, maintenance and diagnosis is arguably the most prominent area after process and planning. The contributions relating to maintenance and diagnosis are somewhat different to the proceeding areas, with theories, models and architectures the most common output from research.

Discussion

RQ1 - What is the publication fora relating to big data in manufacturing?

The rationale behind this research question was to ascertain the level of research interest in the area, as well as highlighting prominent sources of primary research. The results clearly show that big data in manufacturing is an area of research that is experiencing exponential growth, with publications on the topic increasing by a multiple of ten between 2012 and 2014. Looking at the publication results in more depth, there is a correlation between the year-on-year growth in conference and journal publications. This correlation may be a result of early research efforts focusing on the development of short research papers for conferences, and at a later date, developing those papers in to in-depth journal papers. The results identified a number of the most prominent sources of research relating to big data
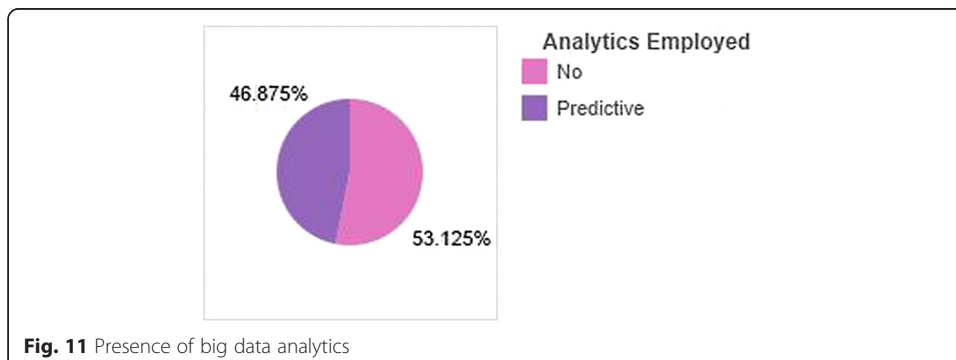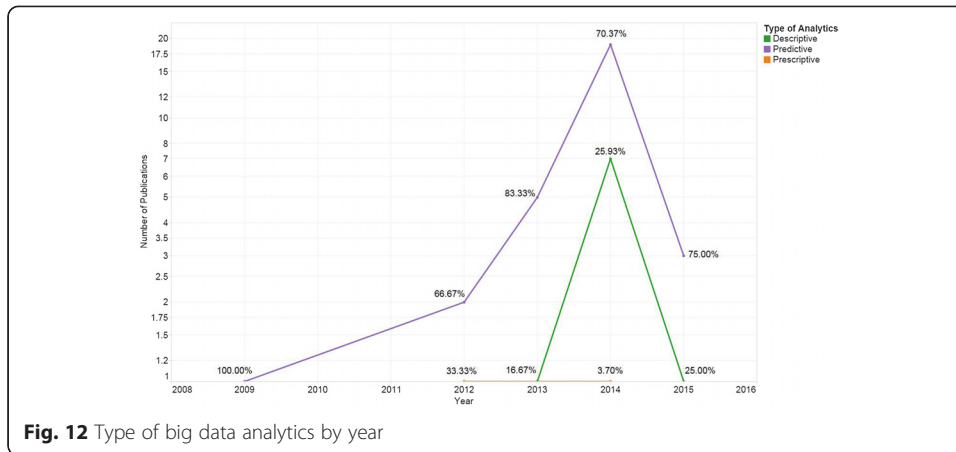

Fig. 11 Presence of big data analytics

Fig. 12 Type of big data analytics by year

technologies in manufacturing, with the Journal of Production Economics, and IEEE Conference on Big Data, being the top sources for journal and conference publications respectively. At present, research interest in the area of big data technologies in manufacturing is high, which is clearly illustrated by the year-on-year exponential growth in publications over the last number of years.

RQ2 - What type of research is being undertaken in the area of big data in manufacturing?

The rationale behind this research question was to understand the maturity of the research area. This was based on the assumption that philosophical research that focuses on theory, with no application or implementation, may be indicative of an area that is relatively immature and requires the development of theory to support future
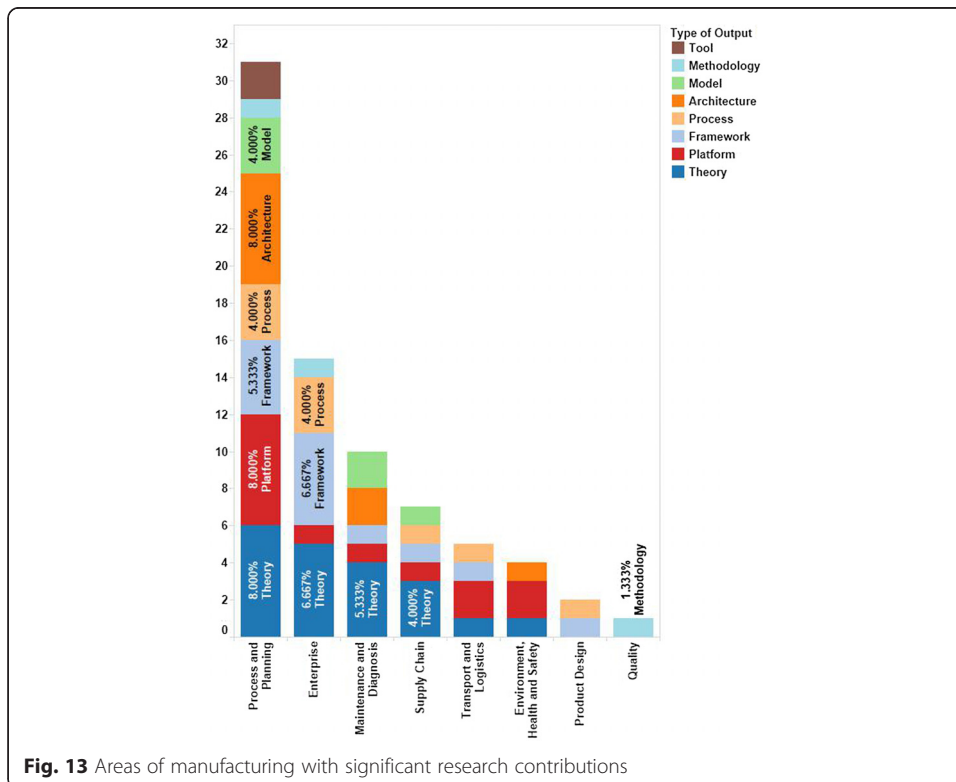

Fig. 13 Areas of manufacturing with significant research contributions

applications. This type of feed-forward process is evident in the results focusing on the type of research approaches being employed, where philosophical research is the most common, with a chronological lag in solution and evaluation research. Most notably, evaluation research, which is considered in this study to be the most mature type of research relating to technology implementations, is associated with the same number of publications in the Q1 2015 as it was in 2013 and 2014 combined. While big data technologies in manufacturing is a new area of research in chronological terms, the exponential growth shown in the results related to RQ1, coupled with the natural cascading of theoretical and philosophical research, with rigorous, empirical and demonstrable research, indicates that the area is developing rapidly. However, based on the findings in this study, the area may still be classified as being somewhat immature due to the high proportion of philosophical-based research, coupled with the low quantity of rigorous evaluation-based research.

RQ3 – What type of contributions are being made to the area of big data in manufacturing?

The rationale behind this research question was to further identify the maturity of the area by classifying the type of outputs originating from research, while also highlighting prominent current research themes and trends. The year-on-year data shows an increasing distribution and balance in the area. This may be interpreted as being indicative of a vibrant research community that is maturing and evolving. There appears to be a strong relationship between results in RQ2 and RQ3. In particular, the most prominent classifications in both sets of results are largely analogous, namely philosophical-based research, and theory-based contributions. As this area of research is relatively new and immature, there is an emphasis on developing theories that can be used by future research efforts to solve particular problems in the field. Indeed, the next most prominent research outputs after theory are frameworks and platforms. These types of outputs can be viewed as a midway point between theory and application, as they are developed on a theoretical foundation (e.g. design or architecture) and facilitate the development of applications and systems. One anomaly in the results showed that there was a lack of journal papers identifying platforms as their output, when compared to that of conferences. This could simply be a result of the term 'platform' being more prominent in one community (e.g. industry conferences), versus another community (e.g. academic journals). However, investigating the anomaly further is not warranted in this study given that it is not critical to answering the research question. As previously alluded to, the main themes in the research contributions overlap to some extent with the findings from RQ2. Based on these results, it appears that there is a theoretical base being developed to progress the research area, with technologies being developed to implement those theories.

RQ4– What type of analytics are being used in the area of big data in manufacturing?

The rationale behind this research question was to identify the extent that analytics is being used with big data technologies in manufacturing, as well as understanding the type of problems being solved. The results show that about half of big data in manufacturing research employs some form of analytics. This is interesting in the sense that it confirms that big data technologies are being used independent of analytics, and the terms should not be used synonymously. The most common type of problems handled by big data

O'Donovan *et al. Journal of Big Data* (2015) 2:20

Page 16 of 22

analytics is prediction accuracy, which is a desirable quality in decision-making. The prominence of predictive analytics may be attributed to the presence of theories and methods pertaining to prediction from other fields (e.g. statistics), and the applicability of prediction analytics to real-world problems. In contrast, the lack of prescriptive analytics is evident from the results. However, this can be attributed to the difficulty in constructing prescriptive applications. Prescriptive applications are inherently complex when compared with descriptive and predictive analytics, given the need to align technology, modelling, prediction, optimisation, and subject matter expertise. Therefore, given the area of big data in manufacturing is still in its infancy, it is little surprise that only a few of these highly complex prescriptive analytics applications have emerged.

RQ5– What areas of manufacturing are big data technologies being applied?

The rationale behind this research question was to highlight research themes relating to big data technologies in manufacturing, with a particular emphasis on understanding the type of manufacturing problems that are being addressed by big data. The results indicate that process and planning in manufacturing is currently receiving the most research interest, with diverse cross-departmental enterprise applications, and maintenance and diagnosis making up 74.6 % of publications in this study. However, given research classified as 'enterprise' is comprised of diverse applications that address a broad range of topics in manufacturing, it should not be classified as a significant research theme for big data in manufacturing. In keeping with the results from RQ2 and RQ3, there is a strong emphasis on theoretical research, as well as the development of frameworks, platforms and architectures to realise those theoretical foundations.

Threats to validity

As with any secondary research methodology, the process of systematic mapping is not infallible, and there are indeed a number of threats to the validity of this study. To this end, every effort was made to mitigate potential risks throughout the process. The threats to the validity that were identified are described in this section.

Search criteria

The search criteria used to acquire papers for this study was chosen collectively by participating researchers. The choices relating to the search criteria were driven by (a) the agreed scope of the research, (b) the research questions that needed to be answered, as well as (c) the relevance of papers returned from testing various search combinations. However, while upmost care was taken to choose the most appropriate search strings for the study, there is an inherent risk that this could restrict the discovery of papers that did not meet the search criteria. Although based on the experience of the research team, the sophistication of the search facilities in modern digital databases, coupled with the availability of publication meta-data, as well as the commonality of the terms incorporated in our search string, would suggest that the risk of omitting relevant papers was at least minimised.

Digital databases

The research team selected the digital databases to acquire papers for the study. These databases were selected using a combination of prior knowledge relating to engineering

and technology research, as well as noting prominent databases used in closely related fields. Therefore, if a particular digital repository was not searched, there is a risk that relevant papers would not be included in the study. However, the amount of overlap experienced in the search results across different types of digital repositories provided a level of redundancy. More to the point, if a particular digital repository was not used in the study, there is a realistic chance that the research which it holds will either be indexed by another source that is being used, or indeed, be discovered by following the references from each papers in the study (e.g. snowballing).

Google scholar

As specified in the research methodology, there was an issue with constructing an appropriate search string for Google Scholar. While other databases enabled the construction of searches to interrogate titles, abstracts and keywords using Boolean logic, Google Scholar was limited to searching by title or full text. When a full text search was carried out a total of 9540 records were returned, which is obviously too much data to analyse for a study of this size. Therefore, the search by title option was chosen as it returned a manageable 14 publications. By choosing this search approach for Google Scholar, there is a risk that publications with abstracts and keywords that match the study's search criteria may have been omitted.

Inclusion and exclusion criteria

The criteria defined for inclusion and exclusion in this study stemmed from discussions within the research team, where the rules and conditions that were deemed to be aligned with the scope of the study were identified. Creating rules to identify the initial literature to review means that there is a risk that relevant research may be omitted if it utilises different terminology to that of the inclusion/exclusion criteria. However, the study's primary search terms, namely manufacturing and big data, are conventional, well-defined and accepted terms, which should reduce the number of publications omitted due to authors using synonymous terms. Furthermore, as the study is focused on identifying the main research in the area of big data in manufacturing, there is not as much of a concern with capturing research that is very loosely related to the topic.

Classification accuracy

There is a risk that the research teams labelling and categorisation of the research in the study may be different to that of another researcher. To reduce individual bias, and gain confidence in the accuracy of our classification process, each researcher in the team was asked to classify each publication. The results of this classification process was then analysed, with those publications that were classified the same being labelled immediately, and those with differing classifications subject to a review meeting to determine the most relevant classification.

Conclusion and future work

At the time of writing, this is the only research effort focusing on the systematic mapping of big data technologies in manufacturing. The research presented in this paper provided a breadth-first review of the research relating to big data in manufacturing to

O'Donovan *et al. Journal of Big Data* (2015) 2:20

Page 18 of 22

promote a better understanding of a new and pervasive area. In particular, several fundamental research questions that are relevant to current research efforts focusing on big data in manufacturing were answered, while also providing an excellent platform for further research and investigation in the area. In particular, it is logical that future work should focus on the development of systematic and literature reviews that are aligned with the areas of manufacturing identified in this study, such as the creation of a systematic review of big data in manufacturing that is focused on maintenance and diagnosis. The combination of these reviews, coupled with the systematic mapping presented in this research, can serve to provide a complete perspective of the primary research relating to big data in manufacturing.

Appendix
Research papers

Table 8

Publication	Title	Year	Classification
IEEE International Conference on Industrial Engineering and Engineering Management	Maintenance behaviour-based prediction system using data mining	2009	Architecture
Journal of Systems Science and Systems Engineering	The next industrial revolution: Integrated services and goods	2012	Theory
Procedia CIRP	An Agent-Based Model to Evaluate Carpooling at Large Manufacturing Plants	2012	Theory
ASME International Mechanical Engineering Congress and Exposition	Cultivating emerging and black swan technologies	2012	Methodology
Journal of the Chinese Institute of Industrial Engineers	Manufacturing intelligence for early warning of key equipment excursion for advanced equipment control in semiconductor manufacturing	2012	Process
The International Society for Optical Engineering	A distributed big data storage and data mining framework for solar-generated electricity quantity forecasting	2012	Platform
ASME International Mechanical Engineering Congress and Exposition	Next generation technologies for improving product planning and development - An industry perspective	2013	Theory
Database Systems Journal	About Big Data and its Challenges and Benefits in Manufacturing	2013	Theory
Lecture Notes in Computer Science	Big data challenges in industrial automation	2013	Theory
Manufacturing Letters	Recent advances and trends in predictive manufacturing systems in big data environment	2013	Theory
AUTOTESTCON	Using big data and predictive machine learning in aerospace test environments	2013	Model
Journal of Business Logistics	Click Here for a Data Scientist: Big Data, Predictive Analytics, and Theory Development in the Era of a Maker Movement Supply Chain	2013	Model
Procedia CIRP	Wavelet based De-noising in Manufacturing and in Business	2013	Model
Annual Symposium on Cloud Computing	Compiling machine learning algorithms with systemML	2013	Platform
Applied Mechanics and Materials	Cloud Service Platform for Big Data of Manufacturing	2013	Platform

O'Donovan *et al. Journal of Big Data* (2015) 2:20

Page 19 of 22

Table 8 *(Continued)*

IEEE 14th International Symposium on "A World of Wireless, Mobile and Multimedia Networks	Enhanced multi-domain access control for secure mobile collaboration through Linked Data cloud in manufacturing	2013	Platform
Advanced Semiconductor Manufacturing Conference	Heading towards big data building a better data warehouse for more data, more speed, and more users	2013	Process
Advanced Semiconductor Manufacturing Conference	Leveraging applied materials TechEdge Prizm™ for advanced lithography process control	2013	Process
IEEE Transactions on Biomedical Engineering	Grand challenge: applying regulatory science and big data to improve medical device innovation	2013	Process
International Conference on Computers and Industrial Engineering	Mining logistics trajectory knowledge from RFID-enabled production big data	2013	Process
IEEE International Conference on Big Data	A system architecture for manufacturing process analysis based on big data and process mining techniques	2014	Architecture
IEEE International Conference on Big Data	Advanced planning and control of manufacturing processes in steel industry through big data analytics: Case study and architecture proposal	2014	Architecture
IEEE International Congress on Big Data	Big Data Analytics for Predictive Manufacturing Control - A Case Study from Process Industry	2014	Architecture
IEEE International Congress on Big Data	Big Data Analytics for Predictive Manufacturing Control - A Case Study from Process Industry	2014	Architecture
Lecture Notes in Computer Science	Energy data analytics towards energy-efficient operations for industrial and commercial consumers	2014	Architecture
Lecture Notes in Informatics	Realizing the predictive enterprise through intelligent process predictions based on big data analytics: A case study and architecture proposal	2014	Architecture
Fall Simulation Interoperability Workshop	PMML in manufacturing applications	2014	Framework
IEEE International Conference on Big Data	Towards a domain-specific framework for predictive analytics in manufacturing	2014	Framework
IEEE International Conference on Big Data	Big automotive data: leveraging large volumes of data for knowledge-driven product development	2014	Framework
IEEE International Conference on System Science and Engineering	Network manufacturing in the big data environment	2014	Framework
IEEE Transactions on Semiconductor Manufacturing	A Framework for Root Cause Detection of Sub-Batch Processing System for Semiconductor Manufacturing Big Data Analytics	2014	Framework
International Conference on Biomedical Engineering and Informatics	A framework for management of massive knowledge in cloud environment	2014	Framework
International Conference on Information Fusion	Prescriptive information fusion	2014	Framework
International Journal of Computational Intelligence Systems	A data mining approach for analyzing semiconductor MES and FDC data to enhance overall usage effectiveness (OUE)	2014	Framework
Key Engineering Materials	Modeling for global management of complex product subject data oriented at big-data	2014	Framework
Applied Mechanics and Materials	Investigation of the Datamation of Manufacturing Industrial Chain in the Big Data Era	2014	Theory

O'Donovan *et al. Journal of Big Data* (2015) 2:20

Page 20 of 22

Table 8 *(Continued)*

Applied Mechanics and Materials	Data flow analysis on 3D printing for distributed manufacturing information system	2014	Theory
IEEE International Conference on Big Data	Big data predictive analtyics for proactive semiconductor equipment maintenance	2014	Theory
IEEE International Parallel & Distributed Processing Symposium	Machine-Learning-Based Identification of Defect Patterns in Semiconductor Wafer Maps: An Overview and Proposal	2014	Theory
IIE Annual Conference and Expo 2014	Integration of reliability, service and renewables for sustainable manufacturing initiatives	2014	Theory
InTech	Industrial big data analytics: The present and future	2014	Theory
Journal of Business Logistics	Supply chain game changers-mega, nano, and virtual trends-and forces that impede supply chain design (i.e., Building a Winning Team)	2014	Theory
Journal of Business Logistics	The supply chain becomes the demand chain	2014	Theory
Procedia CIRP	Service Innovation and Smart Analytics for Industry 4.0 and Big Data Environment	2014	Theory
IEEE International Conference on Mechatronics and Automa	Error correction of optical path component manufacture for Fiber Optic Gyroscope using SVM and Monte Carlo	2014	Methodology
Winter Simulation Conference	Data analytics using simulation for smart manufacturing	2014	Methodology
Applied Stochastic Models in Business and Industry	Modeling and analyzing semiconductor yield with generalized linear mixed models	2014	Model
International Journal of Simulation Source of the Document International Journal of Simulation Modelling	Batch task scheduling-oriented optimization modelling and simulation in cloud manufacturing	2014	Model
IPC APEX EXPO 2014	Enhancing visibility and agility in the electronics manufacturing supply chain	2014	Model
ACM SIGKDD International Conference on Knowledge Discovery and Data Mining	Applying data mining techniques to address critical process optimization needs in advanced manufacturing	2014	Platform
Procedia CIRP	A Solution for Information Management in Logistics Operations of Modern Manufacturing Chains	2014	Platform
IEEE International Conference on Computer Supported Cooperative Work in Design	A big data cleansing approach for n-dimensional RFID-Cuboids	2014	Process
Lecture Notes in Business Information Processing	Developing Data Analytics to Improve Services in a Mechanical Engineering Company	2014	Process
Neural Computing and Applications	Large, high-dimensional, heterogeneous multi-sensor data analysis approach for process yield optimization in polymer film industry	2014	Process
Winter Simulation Conference	Simulation for dedicated line small lot size manufacturing	2014	Process
Winter Simulation Conference	Virtual factory revisited for manufacturing data analytics	2014	Theory
Journal of Digital Convergence	An Exploratory Study on Application Plan of Big Data to Manufacturing Execution System	2014	Theory
Procedia CIRP	Short-term Cyber-physical Production Management	2014	Theory

O'Donovan *et al. Journal of Big Data* (2015) 2:20

Page 21 of 22

Table 8 *(Continued)*

Solid State Device Research Conference	The Factory Integration Roadmap in Semiconductor manufacturing	2014	Theory
ACM SIGKDD International Conference on Knowledge Discovery and Data Mining	Applying data mining techniques to address critical process optimization needs in advanced manufacturing	2014	Platform
Euro Med Telco Conference - From Network Infrastructures to Network Fabric: Revolution at the Edges	Cloud based system for measurement data management in large scale electronic production	2014	Platform
IEEE International Conference on Service Operations and Logistics, and Informatics	ACP approach based intelligent quality management system for manufacturing processes	2014	Platform
IEEE International Conference on Service Operations and Logistics, and Informatics	ACP approach based intelligent quality management system for manufacturing processes	2014	Platform
IEEE World Congress on Services	Cloud Computing for Big Data Entrepreneurship in the Supply Chain: Using SAP HANA for Pharmaceutical Track-and-Trace Analytics	2014	Platform
Procedia CIRP	Predictive Analytics Model for Power Consumption in Manufacturing	2014	Platform
The International Society for Optical Engineering	New techniques in large scale metrology toolset data mining to accelerate integrated chip technology development and increase manufacturing efficiencies	2014	Platform
IEEE International Conference on Big Data	Toward smart manufacturing using decision analytics	2014	Tool
Winter Simulation Conference	Big data in daily manufacturing operations	2014	Tool
International Journal of Production Economics	A big data approach for logistics trajectory discovery from RFID-enabled production data	2015	Architecture
Manufacturing Letters	A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems	2015	Architecture
International Journal of Production Economics	Managing a Big Data project: The case of Ramco Cements Limited	2015	Framework
International Journal of Production Economics	The value of Big Data in servitization	2015	Framework
International Journal of Production Economics	Insights from hashtag #supplychain and Twitter Analytics: Considering Twitter and Twitter data for supply chain practice and research	2015	Framework
Studies in Computational Intelligence	Manufacturing Cyber-Physical Systems Enabled by Complex Event Processing and Big Data Environments: A Framework for Development	2015	Framework
Studies in Computational Intelligence	Manufacturing Operations, Internet of Things, and Big Data: Towards Predictive Manufacturing Systems	2015	Theory

O'Donovan *et al. Journal of Big Data* (2015) 2:20

Page 22 of 22

References

1. Davis J, Edgar T, Porter J, Bernaden J, Sarli M (2012) Smart manufacturing, manufacturing intelligence and demand-dynamic performance. Comput Chem Eng 47:145–156
2. Chand S, Davis J (2010) What is smart manufacturing? Time magazine
3. Lee J, Kao HA, Yang S (2014) Service innovation and smart analytics for Industry 4.0 and big data environment. Procedia CIRP 16:3–8
4. Hazen BT, Boone CA, Ezell JD, Jones-Farmer LA (2014) Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. Int J Prod Econ 154:72–80
5. Fosso Wamba S, Akter S, Edwards A, Chopin G, Gnanzou D. How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study. Int J Prod Econ. 2015;165:1–13.
6. Lee J, Lapira E, Bagheri B, Kao H (2013) Recent advances and trends in predictive manufacturing systems in big data environment. Manuf Lett 1(1):38–41
7. Kumar P, Dhruv B, Rawat S, Rathore VS (2014) Present and future access methodologies of big data. Int J Adv Res Sci Eng 8354(3):541–547
8. McKinsey, "Big data: The next frontier for innovation, competition, and productivity," 2011.
9. Philip Chen CL, Zhang C-Y (2014) "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data,". Inf Sci (Ny) 275:314–347
10. Vera-baquero A, Colomo-palacios R, Molloy O (2014) "Towards a process to guide Big data based decision support systems for business processes,". In: Conference on ENTERprise information systems towards, vol 00., p 2212
11. Lee J, Bagheri B, Kao H (2015) A cyber-physical systems architecture for industry 4. 0-based manufacturing systems. Manuf Lett 3:18–23
12. Wright P (2014) Cyber-physical product manufacturing. Manuf Lett 2(2):49–53
13. Petersen K, Feldt R, Mujtaba S, Mattsson M (2008) "Systematic mapping studies in software engineering," EASE'08 Proc. 12th Int Conf Eval Assess Softw Eng., pp. 68–77.
14. Budgen B, Turner D, Brereton M, Kitchenham P (2008) Using mapping studies in software engineering. Proc PPIG 2:195–204, 2088
15. Wieringa R, Maiden N, Mead N, Rolland C (2006) Requirements engineering paper classification and evaluation criteria: a proposal and a discussion. Requir Eng 11:102–107
16. Meziane F, Vadera S, Kobbacy K, Proudlove N (2000) Intelligent systems in manufacturing: current developments and future prospects. Integr Manuf Syst 11(4):218–238
17. Delen D, Demirkan H (2013) Data, information and analytics as services. Decis Support Syst 55(1):359–363