A Contextual Scientometric Analysis of Indonesian Biomedicine Mapping the Potential of Basic Research Downstreaming

Ria Hardiyati, Irene Muflikh Nadhiroh, Tri Handayani, V.M. Mesnan Silalahi, Rizka Rahmaida, and Mia Amelia

Research Center for Development of Science & Technology, Indonesian Institute of Sciences

JOURNAL OF SCIENCE, TECHNOLOGY, AND INNOVATION POLICY AND MANAGEMENT (STIPM JOURNAL), Volume 03, Number 01, July 2018

FOREWORD by EDITOR-in-CHIEF

We are glad to announce that the journal of *Science, Technology, & Innovation Policy and Management* (STIPM Journal) Vol 3, No. 1, July, 2018 is ready for public reading and views. The journal itself focus on STI policy and management.

The aim of this issue is to combine the various perspectives of R&D management and STI policy. Original papers as well as case studies-based research are presented to the readers.

STIPM Journal is an online research journal managed by the Center for Science and Technology Development Studies, Indonesian Institute of Sciences (PAPPIPTEK-LIPI). This journal is a blind peer reviewed journal, which provides free access to research thoughts, innovation, and original discoveries that are needed mostly by the research scholars. In this edition, the STIPM Journal contains six articles dealing with science, technology and innovation policy and management written by scholars from Japan, Thailand, India and Indonesia.

The first article, entitled *India's science, technology and innovation policy: Choices for course corection with lessons learned from China* by **G.D. Sandhya.** In this paper, an attempt has been made to look at how comprehensive India's STI policies with regard to policy components; a roadmap; and strategies for execution and boldness in terms of identifying and recognising the failures and recommend major structural changes. What is intended is to understand the relationship between the domain of S&T policy and expected outcomes; the mismatch between the policy expectations and outcomes. An attempt is being made to identify possibility for correction by taking lessons from other economies, such as China.

Second article were written by **Wati Hermawati, et al.,** entitled *Outcome and impact based evaluation of research program implementation: A case of Indonesian public research institute.* This article relates to outcome and impact based evaluation (OIBE) of a research program implementation at an Indonesian public research institute (PRI) 'A'. The major funding for PRIs in Indonesia comes from government. It is very essential, therefore, for various parties including policy makers to be informed about meaningful and relevant evaluation of the outcome and impact of such PRI to the welfare of the people, to technology development and innovation, and to the policy improvements in significant ways.

Hidenori Shigeno, et al., presents the third article, *Internal innovation capability and ICT use in the innovation process from the view of connectivity in Japanese SMEs.* This article discusses how internal innovation capability such as the technological level and R&D (Research and Development) contributes to the innovation and how it is promoted by ICT use. Using the survey data of about 650 SMEs (Small Medium Enterprise) from all over Japan, this study constructs two models with ICT or without ICT and focuses on how SEMs (Structural Equation Modeling) obtain information from external linkages and the role of ICT in the innovation process

The effect of team diversity in cross-functional teams for enhancing research commercialization: An experience of Thai public research institute is an article presented by **Warangkana Punyakornwong**. This article discusses the effect of team diversity and institutional factors in terms of top management support and incentive system on the number of license agreements in the context of the National Science and Technology Development Agency (NSTDA) in Thailand.

The fifth article entitled *A contextual scientometric analysis of Indonesian biomedicine: Mapping the potential of basic research downstreaming* is presented by **Ria Hardiyati, et al**. The article discusses how to obtain a rich contextual overview of the development of biomedicine research in Indonesia, for example in the context of the down-streaming potential of research publications. The results of text data processing using a computational model and bibliometric analysis will provide a richer contextual picture as a proxy to reveal the potential for down-streaming of basic research.

Final article was compiled by **Kristiana, et al.,** with the title *The value chain analysis to support industrial cluster development of oil palm-cattle integration in Pelalawan Regency, Indonesia.* This article discusses the value chain of oil palm-cattle integration proggram and to formulate reinforcement programs to develop cluster of oil palm-cattle integration with industrial cluster approaches. Among the five products from the oil palm-cattle integration program, the liquid organic fertilizer and solid manure are more profitable than the primary product of husbandry: the beef. Nonetheless, both products are highly dependent on the beef cattle existence. In other words, if the business of manure and liquid organic fertilizer are not profitable, the business of beef cattle will also fail.

In addition to all articles that presented in this volume, we also would like to thank the authors, editors, and reviewers who have worked very hard in this edition. We hope that all articles featured in this edition will be useful for the reader.

Jakarta, 16 July 2018

Editor-in-Chief

JOURNAL OF STI POLICY AND MANAGEMENT

Volume 3, Number 1, July 2018

LIST OF CONTENTS

STI POLICY AND MANAGEMENT

A Contextual Scientometric Analysis of Indonesian Biomedicine: Mapping the Potential of Basic Research Downstreaming

Ria Hardiyati*, Irene Muflikh Nadhiroh, Tri Handayani, V.M. Mesnan Silalahi, Rizka Rahmaida, and Mia Amelia
Research Center for Development of Science & Technology, Indonesian Institute of Sciences

ARTICLE INFO	ABSTRACT
	A content analysis on Indonesian biomedical research publication is conducted based on text mining. The research is necessary to obtain a rich contextual overview of the development of biomedicine research in Indonesia, for example in the context of the downstreaming potential of research publications. The results of text data processing using a computational model and bibliometric analysis will provide a richer contextual picture as a proxy to reveal the potential for downstreaming of basic research. Quantitative research is conducted using data sources from Scopus, Google Scholar and universities's and scientific journal's repository to analyse the performance at the meso and macro levels. Interpretation of the results is qualitatively carried out within a focus group discussion (FGD) session with domain experts. An attempt is carried out to see the trajectory of series of research publications from basic research stage down to clinical trials. This effort results in an ability to show the research trajectory in anti-malaria drug development from the basic research, which evolves to the clinical trial. This study reveals many discontinuities in the trajectory of a research topic from basic research to downstreaming in drugs development in Indonesia.

I. INTRODUCTION

Indonesia's biodiversity is a great potential for the development of industrial raw materials, one of which is medicine. The potential comes from not only plants, but also animals and microbes, e.g. insect *Aristolochia* sp. for anti-bacteria, fungi from class Actinomycetes for anti-tumor (Sukara & Tobing, 2008). However, there are only few biomedicine researches that can be applied in industries or reach research downstream.

Until now, it is mentioned in Decree of Minister of Health No. 381/2017 that there are around 9,600 types of medicinal plants that can be utilized for drug development. With such enormous potential, biomedicine research in Indonesia should have been developed quite drastically. Meanwhile, data on Indonesian scientific publications in the domain of biomedicine has grown in significant amounts. This data can be used as materials for analysis to get a picture

* Corresponding Author. +628122613025
E-mail: ria.hardiyati@yahoo.co.id

of the development of science and technology in biomedicine in Indonesia. Previous studies to analyze biomedicine publications in Indonesia have been conducted in the form of bibliometric studies and co-authorship network analysis (Handayani, Hardiyati, Amelia, Rahmaida, & Nadhiroh, 2016). However, rich information stored in text content has so far not been analyzed. To this end, a scientometric study is conducted in the analysis of scientific publication content in biomedicine. The objectives of this research are to develop a potential downstreaming classification model of scientific publications in the field of biomedicine based on text mining, and to analyze the downstreaming potential of scientific publications in Indonesia. To this end, a combined bibliometric and scientometric analysis is conducted to get a richer contextual information of the research publications study.

A. Text-mining Based Scientometrics

In this paper, bibliometrics is defined as "the application of mathematics and statistical methods to books and other media of communication" (Pritchard, 1969) and scientometrics is defined as "the quantitative methods of the research on the development of science as an informational process" (Nalimov & Mulcjenko, 1971). Both of this study often combined together for the rich results of a content analysis on research publications. Scientometrics study attempts to measure the evolution of a scientific domain, the impact of scholarly publications, the patterns of authorship, and the process of scientific knowledge production. Scientometrics is also an evolving field of study which is influenced by the recent advancement of technology. New definition is developed to accomodate a broader context in scientometrics, such as that, investigated by Bornmann and Haunschild (2017) for the new orientation towards societal impact. Scientometrics and bibliometrics often involve the monitoring of research, the assessment of the scientific contribution of authors, journals or specific works, as well as the analysis of the dissemination process of scientific knowledge. Many contextual scientometrics studies were developed due to limitation of the bibliometrics

study output. Due to lacking of intra-institutional systematic research quality and trend evaluation models, Hassanain, Anil, and Abdo (2016) developed a conceptual framework to evaluate the trends and impact of institutional research. In order to know an index to measure the productivity of author or journal with respect to a particular research context, Lathabai, Prabhakaran, and Changat (2017) developed an approach to devise a network scientometrics to address the identification of contextual productivity.

Text-mining based scientometrics study is mainly carried out based on a text corpus of research publications. In this context, many researches are directed to discover new knowledge in making intelligent decision. Luo, Wu, Gopukumar, and Zhao (2016) reviewed and discussed big data application in four major biomedical subdisciplines, whereby rich output experiments using big data facilitate the new studies of diseases. Text-mining based scientometrics study is still lacking in Indonesia. This study is important because previous studies have been done mostly based on bibliometrics analysis. The bibliometrics analysis is based on counting the quantitative bibliographical data which is not rich in the context and description of the results. Meanwhile, the research publication data available on the internet is more easily accessible and exists in a greater volume, which is often called big data. To optimize these resources, it is necessary to conduct a scientometrics study based on text mining. When viewed from the benefits of specific text-based research in scientometrics study in biomedicine in Indonesia, the results of this study can provide richer analysis in describing the trend of technological development in this domain in general.

The development of information and computational technology has made it possible to conduct scientometrics studies contextually using the large amount of available research publication data, obtained online and using natural language processing techniques and a variety of increasingly sophisticated big data computation algorithms. Therefore, a scientometrics study on this basis provides the ability to present the results of richer and contextual data analysis information in new ways. This article presents the results of

a study on Indonesian biomedicine publications using text mining in discovering new knowledge in the development of Indonesian biomedicine research.

B. Research Downstreaming Process

The development of a new drug requires a long time and high cost research. Research that produces a new drug begins with basic studies that generate thousands of compounds. According to the US Food and Drug Administration (FDA), the stages of biomedicine R&D process consists of the discovery or concept stage (learning about disease, generating thousands compound potential, and conducting a series of tests to obtain promising lead compound), pre-clinical research (in vitro and in vivo), clinical research (consisting of 3 phases, tested to humans), FDA review, and FDA post-market safety within months or years. Meanwhile, according to the Regulation of Head of National Agency of Drug and Food Control (BPOM) No. 16 Year 2015, as the regulator of a drug distribution permit, the phases of drug development process consists of a drug discovery phase, non-clinical test (in vitro and in vivo pharmacology, toxicity and pre-clinical test), and a 4 phases clinical development Meanwhile, according to The Pharmaceutical Research and Manufacturers of America (PhRMA), a bio-pharmaceutical research company in the United States, the stages of R&D process consists of basic research (studying the disease and generate potential compounds), drug discovery (do some tests to generate promising or lead compound), pre-clinical (in vitro and in vivo), clinical trials (Phase I-III), FDA review, clinical trial phase IV or post-approval research & monitoring. These stages are also similar to that described by Hughes, Rees, Kalindjian, and Philpott (2011) which consist of basic research, lead-compound discovery, pre-clinical development, clinical development, FDA filing. Meanwhile, according to PT. BioFarma, the stages consist of vaccine research, pre-clinical testing, phase I clinical trials, phase II clinical trials, phase III clinical trials, and phase IV clinical trials.

The main goal of the study is to be able to do a technology readiness assessment. This is important in order to reduce risk in budget and planning as well as in focusing on the down-streaming of research results. One of the tools to help evaluate process of research downstream is technology readiness levels (TRL) scale. TRL scale was developed by NASA in 1970–1980's. NASA developed the first 7 levels which was further developed into 9 levels. Level 9 means that the technologies had gained acceptance from industries and government (EARTO, 2014). Adaption of the TRL approach can be found in the US-Department of Health and Human Services, NATO, and the US-Departement of Defence. In 2016, the Ministry of Research, Technology and Higher Education of Indonesia issued Ministerial Regulation No. 42 Year 2016 on Measurement and Determination of technology readiness level (TRL). Measurement and determination of the level of technological readiness should be done to determine the preparedness of a technology and reduce the risk of failure in technology utilization. TRL is the level of maturity or readiness of a particular technological research and development result that is measured systematically in order to be adopted by the user, either by government, industry or society (Figure 1).

TRL indicates the level of preparedness of research and development results which will then enter the commercialization stage, so that it can be utilized by the government, industry, and society. TRL consists of 9 levels with different indicators at each level. Researches at levels 1–3 start from basic research, level 4–6 go into applied research, and level 7–9 finally enter into development research. R&D in biomedicine goes into the pharmaceutical category of health. The TRL based text classification will require much
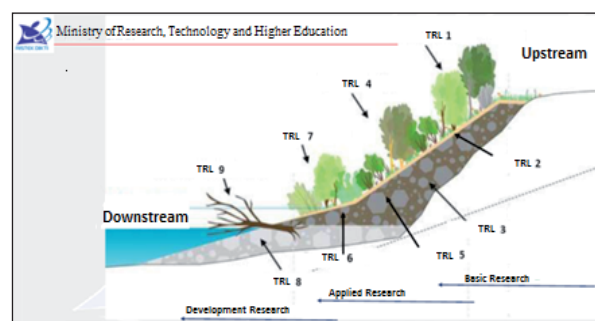

Figure 1. The Research Downstreaming

complex computation if we want to apply in the text mining of research publication. Therefore, we develop a much simpler way to research publication classification as a proxy to show its potential to research downstreaming.

Merriam Webster states that "Definition of Downstream is (1) in the direction of the mouth of a stream or (2) in or toward the latter stages of a industrial production or the stages (such as marketing) after manufacture" (Biomedicine, 2018). Meanwhile the definition of downstream is "relating to or happening at a later stage in a process" (Downstream, 2018). Minister of Research, Technology and Higher Education, Mohamad Natsir, explained that the downstreaming is the implementation of research methods in various ways, so that the community can also enjoy the results (Rachman, 2015). For example, benefit-oriented research, such as ways to prevent flooding or research that can be applied in the industrial world. Meanwhile according to Radjasa on USAID (2015), the results of research or research activities should not just stop being a report, published, and patented, but it should be made a commercialized product for the welfare of the community and for the downstreaming of research results. It is not necessarily in large industries only, but also includes downstreaming in small industries.

In this paper, research results downstreaming is the implementation of research or research results so that it can easily enter the market to be commercialized and or beneficial to the public. In general, there is no standard definition of downstreaming concept in Indonesia. For the purposes of this study, the downstreaming concept is proxied by assuming that the further the publication of the research results in the drug design process, the higher the potential for downstreaming of the results of the study. (Silalahi et al., 2018b). The document classification model is divided into four stages, where the first phase of basic research is seen as a small potential still far from downstream, while the clinical trial stage is seen as to have the greatest potential for downstreaming. This text mining-based document classification model is therefore used as a proxy to see the potential of research results based on its publication. Therefore, to obtain more accurate information related to this potential, a deeper survey is needed, especially for publications claiming to have reached the pre-clinical (PC) and clinical trial stage. To this end, we conducted several surveys based on samples of publication data that have been identified to reach clinical trial (CT) stage in order to get the facts from the field.

Based on the results of literature studies and in-depth interviews with experts related to the stages of biomedicine R&D process, the stages of biomedicine R&D process used in this research consist of four stages: basic research, drug discovery, pre-clinical, and clinical trial. In this sub-section, in addition to providing a brief description for each stage, it also provides a link between the stage and the TRL to determine the potential for biomedicine research downstreaming.

II.   FRAMEWORK OF CONTEXTUAL SCIENTOMETRICS ANALYSIS

Scientometrics studies may focus on contextual information such as how concepts and research topics are being defined over time or in different domains. Many researchers have analyzed large corpora consisting of tens of thousands of journal abstracts to have a rich picture and contextual information on the S&T development in a country or globally. Contextual analysis can be executed using various methods. Using co-word analysis and hierarchical clustering, Stephen (1999) and West (2007) identified relationships among concepts as well as changes in topics studied over time. Abstracts are one of the good indicators to understand and grasp the content of publications. They often provide a more detailed description of the scope of a paper, of its theoretical underpinning, its methodological approach and of other contextual variables. Keywords may also be good indicators of a paper's content, and have often been used in scientometrics studies.

A. Document Classification

The analysis technique used in this study is text mining based on algorithms using machine learning. Essentially, tokenization of the unstructured

raw data (bag of words) will transforms the data into a recognizable format, not only by human users as a data table, but most importantly by machine learning algorithms that require such training data. This data is called a document vector or term document matrix (TDM) and is the cornerstone of the pre-processing necessary for text mining. In the pre-processing the stop-words are removed and stemming is applied. After that, the word-weight filtering is carried out in the corpus through the calculation of TF-IDF. Then, a computational model is built through the comparison of various algorithms in the supervised learning methods that are often used in document classification such as k-Nearest Neighbors (kNN), Naive Bayes (NB), and Super Vector Machine (Cortes & Vapnik, 1995; Joachims, 1998). The labeling of the training and testing dataset can be carried out by the domain experts. In this study, the basic assumption is that the text mining-based document classification model can be used as a proxy to see the potential of research downstreaming. The conceptual model for the document classification model is divided into four stages: basic research (BR), drug development (DD), pre-clinical (PC) and clinical trial (CT). The first phase i.e. basic research (BR) is seen as having a small potential since it is still far from downstreaming, while the clinical trial (CT) stage is seen as having the greatest potential for downstreaming.

B. Social Networking Analysis

A network is a representation of a system in which elements (or nodes) are linked by bonds. Most networks are defined as one mode network with a set of nodes similar to each other. However, some networks are actually two-mode networks also known as bipartite networks. This network is of a certain type, with two different nodes, and the bonds exist only between the nodes belonging to different circuits. Differences are often made between two groups of nodes based on which group is considered more responsible for bonding than others (Opsahl, 2013). An example of a category of two-mode networking that has become popular in recent years is the scientific collaboration network (Newman, 2001) whereby

the nodes are authors and articles. An author is connected to an article if she/he is listed as a writer in the article. The two mode networks are rarely analyzed without changing the network. Most network sizes are defined solely for single mode networks, and few have been redefined for two-mode networks (Latapy, Magnien, & Del Vecchio, 2008). Converting a two-mode network to a single mode network is carried out using a method known as projection. This method operates by selecting one of two groups of nodes and connecting the nodes of the set when connected to at least one common node in the other group. Although the two mode structures are discarded in this process, it is possible to determine the bonding weights based on the two modes. In particular, bonding weights are often defined as the number of common nodes. This method is extended by Newman (2001) who argues that the tie weight among authors in a network of scholarly collaborations should be ignored if the authors collaborate in articles with many articles.

C. Co-Word Analysis

Co-word analysis uses the most important word/keyword of an article to study conceptual structures on a plane (Callon, Courtial, Turner, & Bauin, 1983). Co-word analysis is a content analysis technique using the co-existence pattern of multiple item pairs (i.e. words or noun phrases) in a collection of texts to identify the relationship between the ideas within the subject area presented in the text. Indexes based on frequency of joint events, such as inclusion index and proximity index, are used to measure the strength of relationships between items. Based on this index, items are grouped into several groups and displayed in a network map (Niknia & Mirtaheri, 2015). Since the introduction of co-word analysis by Callon et. al (1983), this analysis was used to explore the intellectual structure (Bansard et.al, 2007), analyzing research trends (Rebholz-Schuhman et.al, 2007), produced hypotheses and discovered knowledge (Stegmann & Grohmann, 2005). The analysis of co-occurrences using statistical techniques such as hierarchical clustering, multi dimensional scaling as well as visualization tools such as the proximity plot, allow one to promptly identify topics and themes in a discipline.

III. METHODOLOGY

The classification model for the downstreaming was built within a five step processes; starting from conceptual modelling, data searching and selection, data cleaning and pre-processing, models building and selection, and finally evaluation and results interpretation (Silalahi et al., 2018a).

The first step taken at the stage of data collection was to determine the search terms used by conducting literature studies. The data collected in this study was an international publication related to biomedicine in Indonesia derived from various databases such as Scopus, DBLP, PubMed Central, Google Scholar, college repositories, and the like. The type of publication used in this study was limited to articles produced by Indonesian authors. An article was said to be the result of an Indonesian writer if at least one of the author's institutions is located in Indonesia. The results of the search through the Scopus database showed that the number of articles related to Indonesian biomedicine was still quite small. Therefore, additional data for Indonesian biomedicine articles was captured through Google Scholar, university's repository, and the likes. Data selection was carried out to ensure the international scientific publication that was encompassed by various articles, derived from Indonesian writers, and related to Indonesian biomedicine.

After data selection as executed, the next step was data integration. In this step, all data was aggregated to be checked for data duplication. The procedure will ensure data duplication did not occur by identifying the unique identity of each article. If a unique identity of the article was recorded more than once, then the data was reduced so that there was only one unique identity for each article. Based on the results of data integration, we obtained data of international scientific publication biomedicine Indonesia, as many as 539 articles. The evaluation stage uses results from a combination of the data processing and analysis through text-mining with bibliometrics analysis. The evaluation was conducted through an FGD involving three experts in the field of biomedicine in Indonesia.

IV. RESULTS AND DISCUSSION

In this section, we describe the results based on the document classification model on the potential downstream by combining the results obtained from the results of the bibliometrics analysis. The document classification is conducted in order to provide a richer contextual analysis in describing technological development trends in biomedicine in Indonesia. Specifically, it described the result in tracing the research progresses into downstreaming in a particular topic by looking at the development of a research topic by using a keyword-based co-word analysis. The analysis was also aimed at seeing the development of research in the domain of biomedicine at the level of an institution, in this case Indonesian Institute of Sciences (LIPI).

A. Description of International Publication Development of Biomedicine Research Indonesia

The development of a number of publications of Indonesian biomedicine research between 1981 and 2017 can be seen in Figure 2. The figure shows that the number of Indonesian biomedicine research publications tends to increase each year, except in 2015 and 2016 which suffered a sharp decline. The significant number of published publications occurred in 2014, which was 8.16% from the previous year.

The development of categorical publication number of Indonesian biomedicine research based on downstreaming potential can be seen in Figure 3. The picture shows that most publica-
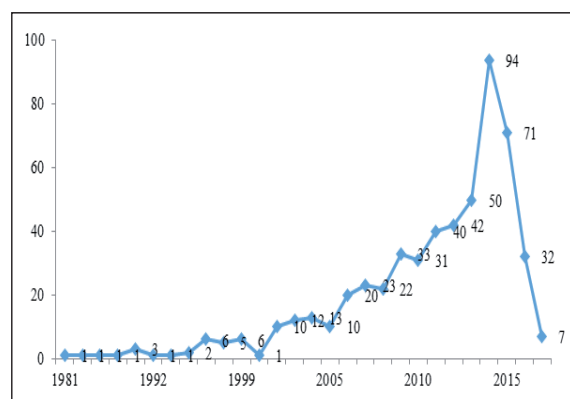
Figure 2. Distribution of Indonesian biomedicine research publication between 1981 and 2017
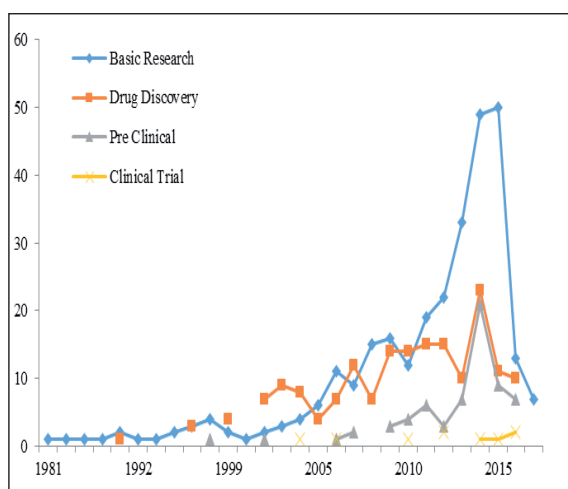
Figure 3. Distribution of Indonesian biomedicine research publication based on drug research and development category between 1981 and 2017

Figure 4. Distribution of Indonesian biomedicine research publication based on drug research and development category (Silalahi et al., 2018b)

tions are in the category of basic research where the largest number of publication growth occurs in 2014–2015. The next dominant category is drug discovery and then pre-clinical category. In the clinical trial category, the research publication start in 2014 and the growth of the number of publications occurred between 2014 and 2016 was very small. This is due to the fact that when the stage of the drug development process reached this stage, the researches were mostly collaborating with the industry which were committed to keep the information secret. This could be mitigated if the government put more funding for the researches at this stage.

Figure 4 shows the distribution of Indonesian biomedicine research publications based on drug research and development categories. Of 539 articles, there were 291 articles (53.99%) of biomedicine research at the basic research stage, 174 articles (32.28%) in drug discovery stage, 65 articles (12.06%) at pre-clinical stage, and 9 articles (1.67%) at the clinical trial stage. This shows that most of Indonesia's biomedicine research is still at a basic research stage where researches are still limited to disease identification; finding potential compound candidates for certain diseases; examining mRNA/protein, genetic associations and phenotypic screening; target identification and validation phase.

Table 1 shows the twenty major institutions with the largest number of biomedical research

publications. The Indonesian Institute of Sciences (LIPI) is an institution with the largest number of publications (125 articles) in biomedicine research, 88 articles from Gajah Mada University (UGM), and 76 articles from Bandung Institute of Technology (ITB). In the context of international collaboration, the data shows big influence of Japan in biomedicine research in Indonesia. Upon searching, from the Scopus database alone it has been identified that Japan has the largest contribution, whereby sequentially the five other countries are Korea, Malaysia, Austria, China, and Netherland. If categorized into phases of drug development trajectory, then the largest number of publications at the basic research and drug discovery stage is owned by LIPI, but at pre-clinical stage the largest publications came from UGM and ITB. Meanwhile, the largest number of publications at the clinical trial stage comes from Dexa Medica Group (industry). The low number of publications at the clinical trial stage in government and university R & D institutions is due to low research funds supporting the later stage of the drug development process which requires a large amount of funding. The lack of facilities and infrastructure (less sophisticated and inadequate laboratory equipment), long bureaucracy, and lack of harmonization and cooperation between government, universities and the industry, and most importantly the unsupported law and regulation are also identified as the factor that also affect the performance of the Indonesian researchers in biomedicine.

Table 1.
Twenty major institutions with the largest number of downstream potential articles of the biomedicine field, 1981–2017

Institution	BR	DD	PC	CT	#
Indonesian Institute of Sciences	64	58	3		125
Gadjah Mada University	45	23	19	1	88
Bandung Institute of Technology	42	17	16	1	76
University of Indonesia	33	21	2	2	58
Airlangga University	23	11	5		39
Bogor Agricultural University	22	9	3		34
University of Illinois at Chicago	6	25			31
Padjadjaran University	13	6	4	1	24
Dexa Medica Group (Indonesia)	3	7	5	5	20
Sam Ratulangi University	5	13			18
Brawijaya University	12	2	3		17
Pancasila University	10	5	1		16
Ehime University	9	3	3		15
Ohio State University	1	13			14
Tohoku Pharmaceutical University	3	9			12
Hasanuddin University	7	4	1		12
University of Amsterdam	2	9			11
Diponegoro University	9	1	1		11
University of Tokyo	4	7			11
Mulawarman University	8	2			10

Figure 5. Co-word visualization at the pre-clinical stage

Figure 6. Co-word visualization at the clinical trial stage

Figure 5 shows the co-word visualization of Indonesian biomedicine research at the pre-clinical stage. The image shows the keyword "*animal experiment, plant extract, in vitro study, in vivo study, drug potency, medicinal plant, non human, animal cell, flavonoid, rat*" are the top ten keywords that are most frequently used terms at this stage. In the pre-clinical phase, in vitro tests (experiments conducted in the laboratory using animal cells showing quality, safety and efficacy) and in vitro tests (experiments performed using animal bodies such as rat showing efficacy, toxicity test, and safety).

The co-word analysis in a particular stage of the development of drugs is attempted to discover new knowledge on the research of Indonesian biomedicine. Figure 6 shows the co-word visual-

ization of Indonesian biomedicine research at the clinical trial stage. The image shows the keyword "human" is the central keyword at this stage. At the clinical trial stage, tests are conducted on the developed compounds or drugs into the human body. High standards are required before testing into the human body which includes candidate compounds or drugs have previously passed toxicity test for at least two types of living things. The co-word analysis reveals that much of the research in this stage is on drug development of anti-malaria and anti-dengue virus. This is in line with the Indonesian government policy on priority setting on the development of drugs for anti-malaria and anti-dengue virus.

Of all the "anti-" related keywords, the "antimalarial" keyword is one of the few keywords that has evolved to the clinical trial stage. Figure 7 shows the co-word analysis of the "antimalarial" keyword at the basic reseach level. It appears that there are several names of

plant species emerging, in addition to the name of the malaria-related parasite species. *Carica papaya* and *Aistonia macrophylia* are two species that emerged from the "antimalarial" keyword group. Both species are believed to contain antimalarial. *Aistonia macrophylia* in addition to being associated with the word antimalarial is also associated with "aorta constriction" which is one of the conditions associated with malaria. In addition there are also keywords "catechin" which is a group of secondary metabolites that are naturally produced by plants.

Figure 8 shows the results of co-word analysis at drug discovery stage for the keyword "antimalarial". It appears that the keywords that appear have been a lot of compound names. *Caesalpinia pulcherrima* is the name of the species that appears at this stage. *Caesalpinia pulcherrima* is a kind of flower plant that is quite popular in Asia, in Indonesia itself commonly referred to as the peacock flower.
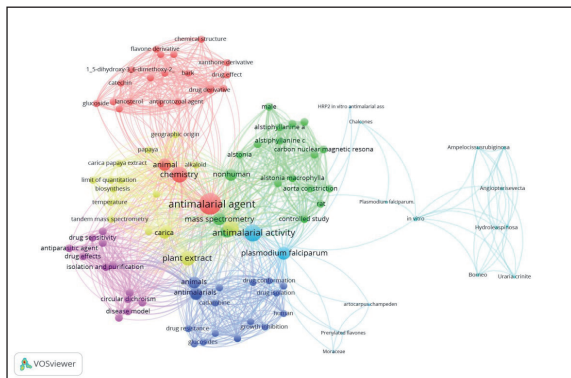

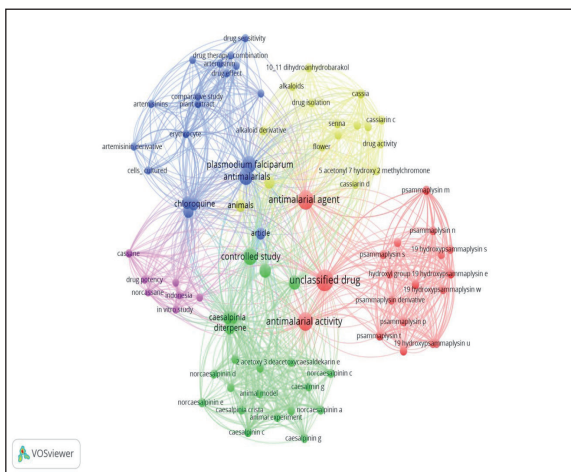**Figure 7.** *Co-word keyword "antimalarial" at the level of basic research*

Figure 8. *Co-word* keyword "antimalaria" at level *drug discovery*

Figure 9 shows the results of co-word analysis on the keyword "antimalarial" at the pre-clinical and clinical trial. The visualization shows that the keywords that appear have been less because there are only a few articles at this stage. Keywords in Figure 9a are associated with activities performed at the pre-clinical stage, such as "in vivo study", "drug response", "dose response", etc, while Figure 9b shows the results of co-word analysis for clinical trial level whereby the keyword "human" is identified which constitute an important discriminatory term to differentiate this category with the others.

Figure 10 shows the results of co-word analysis on articles produced by LIPI authors, related to the "anti-" keyword at the basic research level. It appears that "plant extract" is the most common keyword. In the keyword "plant extract" it also shows some potential drugs derived from plants from LIPI research, such as *Eugenia polyantha* with "antidiabetic", *Achillea milleforium* associated with "antiprotozoa".

Figure 11 shows the results of co-word analysis on articles produced by LIPI authors, related to the "anti-" keyword at the drug discovery level. At the drug discovery level, the results of the LIPI study related to the keyword "anti-" appear to have started specifics.

Figure 12 shows the results of co-word analysis on articles produced by LIPI authors, related to the "anti-" keyword at the pre-clinical level. At the pre-clinical level, the results of the researches in LIPI related to the keyword "anti-" look very specific such as "antidiabetic,
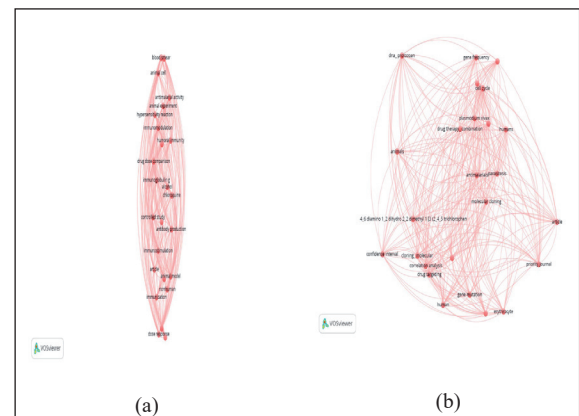

Figure 9. *Co-word* keyword "antimalaria" at level (a) *pre-clinical* and (b) *clinical trial*

Figure 10. *Co-word keywords in the "anti" group often appear in LIPI research at the basic research level*

Figure 11. *Co-word keywords in the "anti" group often appear in LIPI research at the drug discovery level*

Figure 12. *Co-word visualization in LIPI-generated articles is related to the "anti-" keyword at the pre-clinical level*

Figure 13. *Co-word visualization two-mode 6 major institutions with keyword anti -...*

alloxan, hyperglycaemic, kalanchoe pinnata, pancreatic". There is a potency of drug that comes from medicinal plant investigated in LIPI such as *Kalanchoe pinnata*. This medicinal plant is investigated in LIPI as one potential remedy for diabetes (antidiabetic). This plant also contains substances that can inhibit the growth of cancer cells (anticancer).

This chapter also tries to see the topics discussed by six institutions with the most number of articles, namely LIPI, ITB, UGM,

Airlangga University, and Dexa Medika. From the co-word network in Figure 13, it can be seen that the six institutions had included discussions about anti-cancer mix, anti-oxidant, anti-fungal, anti-bacterial, and anti-malaria. In addition, each institution also discusses certain topics, e.g LIPI discusses antidiarrheal and *Scieroderma aurantium* whereby Airlangga University has dicussed topic of antivirus in general.

B. Obstacles in the downstream of Indonesian biomedicine research results, especially Indonesian Institute of Sciences (LIPI)

Some of the explanation for the results of the quantitative analysis are as follows. Based on the results of interviews and FGDs with biomedicine

experts, the obstacles in the downstreaming of Indonesian biomedicine research, particularly in LIPI are as follwes.

First, the search for drug candidates in developed countries has generally been conducted for eight years. In Indonesia, the average focus on developing drug candidates has taken approximately 15 years and requires enormous expenses. For example, Phrma (2015), one of the biopharmaceutical research companies in the United States, stated that the average cost required in drug discovery's research and development is $2.6 billion. This number included the cost of failure, as well as hundreds or even thousands of compounds that may be screened and assessed in the early R&D process. In addition, researchers in Indonesia, especially researchers in LIPI are required to produce research output in the form of national and international journals every year so that researchers can maintain a functional career. Along with this, researchers are also required to be able to produce products from his research. This is what makes Indonesian researchers especially LIPI's researchers overwhelmed to focus in research that leads to the research downstreaming.

Second, LIPI has not yet officially set the direction, whether it is product oriented or science development. Clear regulation within LIPI's operation is very important and needed because when it leads to product oriented and when the product has reached the market; it must generate benefits for researchers and institution. Regulations on the commercialization of research results have actually been issued by the Ministry of Finance but until now there are no technical guidelines.

Third, the problems that often occured and experienced by drug candidate inventors in government R&D agencies and university are those with the rules issued by BPOM. BPOM only accepts the registration of drug candidates from submission of the pharmaceutical industry. To this end, the inventors in government R & D bodies and universities should cooperate with the industry in conducting clinical trials and registration of new drugs. This administrative process is considered troublesome by the inventors. The proposed solution for case of LIPI is that LIPI should create a business entity along with its regulation, where one of the functions of the entity is to register the compound or drug of the research result to BPOM. LIPI's researcher or inventor does not have to bother to cooperate with the pharmaceutical industry with all the complex administrative rules.

Fourth, the pharmaceutical industry is one of the research-based industries whose products are highly regulated, especially in terms of product quality. The pharmaceutical industry has special requirements in the quality management of its products that must meet the rules which is known as the Current Good Manufacturing Practice (cGMP) based on Decree of Minister of Health No. 43 Year 1988. R&D institutions (other than pharmaceutical industries) that have not applied the standard of CPOB will have different quality that is expected by the industry. This will often create misperception in the determination of drug standards from the universities or government R&D agencies with the pharmaceutical industry so that the road to cooperation has been lengthy. In addition, the research project is encouraged to engage the industry, but the industry will not involve if it has not seen the prospects of the research (industry interested if the research results lead at the minimum a level of TRL 7). Drug research, especially herbal medicine in Indonesia is currently in the stage of Technology Readiness Level (TRL) 3 (experimental proof of concept) or 4 (technology validated in laboratory). This condition necessitates the role of government (Ministry of Health and the Central Government) as a liaison between researchers in universities and government R&D agencies with the pharmaceutical industry in fostering the research downstreaming.

V. CONCLUSION

Based on the results obtained, it can be concluded, among others: (1) the downstream potential of a research could be proxied by publication document classification in a drug development process; (2) the results of the document classification are useful in analyzing the development of Indonesian biomedicine research in general and the position of LIPI in this issue; (3) the

results of this research can be extended to make recommendations for Indonesian biomedicine research policy, especially LIPI. If seen from the development trend of scientific publications of biodiversity of Indonesia, it can be concluded that the number of articles on biomedicine research in Indonesia tend to increase from year 1981 until year 2014. This research relies heavily on the feedback of the good experts in constructing conceptual models, labeling, validating and interpreting results. In order to reduce the dependency on domain experts, much larger data is required to enable the machine to do the learning. In this study the amount of data that can be collected is limited. The results require a clear definition of limitations and scope to avoid conflict of interest.

At the meso level, this study gives a picture of the position of an institution, LIPI in Indonesia. This can be seen from the description of research into the downstreaming in a particular topic by looking at the development of research by using a keyword-based co-word analysis. The analysis to see the development trajectory of research topic in biomedicine in LIPI is shown in this study. This study shows the ability to trace the development of research toward downstreaming in a particular topic. The small number of collected publications constraints the results to show the development trajectory of a particular research topic from the basic research stage to the clinical trial stage at institutional level. Therefore, the research needs to be continued to be able to see more comprehensive and detailed development of research in the field of biomedicine in Indonesia, both in terms of text-based data analysis in terms of the amount of data collected or evidences in the field.

At the national level, Indonesia has shifted its focus to the establishment of translational medicine centers, which aims at breakthrough of bottlenecks between basic medical research, clinical research and drug development by translating laboratory findings into clinical applications. Five National Centers for Translational Medicine are under construction now, which will greatly promote the development of medical level and drug development in Indonesia.

REFERENCES

Bansard, J. Y., Rebholz-Schuhmann, D., Cameron, G., Clark, D., van Mulligen, E., Beltrame, E., … Coatrieux, J. L. . (2007). Medical informatics and bioinformatics: A bibliometric study. *IEEE Transaction on Information Technology in Biomedicine*, *11*(3), pp. 237–243.

Biomedicine. (2018) In *Merriam-Webster online dictionary*. Retrieved from https://www.merriam-webster.com/dictionary/biomedicine

Bornmann, L., & Haunschild, R. (2017). Does evaluative scientometrics lose its main focus on scientific quality by the new orientation towards societal impact? *Scientometrics*, *110*(2), pp. 937–943.

(https://doi.org/10.1007/s11192-016-2200-2)

Callon, M., Courtial, J.P., Turner, W.A., & Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information*, *22*(2), pp. 191–235.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*(3), pp. 273–297.

Downstream. (2018) In *Merriam-Webster online dictionary*. Retrieved from https://www.merriam-webster.com/dictionary/downstream

EARTO. (2014). *The TRL Scale as a Research & Innovation Policy Tool*. (EARTO Recommendation Paper). Brussels: EARTO. Retrieved from: http://www.earto.eu/fileadmin/content/03_Publications/The_TRL_Scale_as_a_R_I_Policy_Tool_-_EARTO_Recommendations_-_Final.pdf.

Handayani, T., Hardiyati, R., Amelia, M., Rahmaida, R., & Nadhiroh, IM. (2016). *Kajian saintometrika perkembangan publikasi internasional keanekaragaman hayati Indonesia sebagai bahan rekomendasi kebijakan arah penelitian keanekaragaman hayati nasional*. (Research Paper). Jakarta: Pusat Penelitian Perkembangan Iptek - LIPI.

Hassanain, M., Anil, S., & Abdo, A. (2016). Institutional research evaluation model (IREM): A framework for measuring organizational research trends and impact and its application in medical academia in Saudi Arabia. *Journal of epidemiology and global health*, *6*(4), pp. 249–256.

Hughes J P, Rees S S, Kalindjian S B and Philpott K L. (2011). Principles of early drug discovery Br. *J. Pharmacol, 162*, pp. 1239–1249.

Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. *Machine learning, ECML-98*, pp. 137–142.

Latapy, M., Magnien, C., & Del Vecchio, N. (2008). Basic notions for the analysis of large two-mode networks. *Social networks, 30*(1), pp. 31–48.

Lathabai, H. H., Prabhakaran, T., & Changat, M. (2017). Contextual productivity assessment of authors and journals: A network scientometric approach. *Scientometrics, 110*(2), pp. 711–737. (https://doi.org/10.1007/s11192-016-2202-0)

Luo, J., Wu, M., Gopukumar, D., & Zhao, Y. (2016). Big data application in biomedical research and health care: A literature review. *Biomedical Informatics Insights, 8*, pp. 1–10. (https://doi.org/10.4137/BII.S31559)

Nalimov, V., & Mulcjenko, B. (1971). *Measurement of science: Study of the development of science as an information process*. Washington DC: Foreign Technology Division.

Rachman, T. (2015). Hilirisasi penelitian harus bermanfaat bagi masyarakat. *Republika*. Retrieved from https://www.republika.co.id/berita/pendidikan/dunia-kampus/15/09/23/nv4wdj219-hilirisasi-penelitian-harus-bermanfaat-bagi-masyarakat

Newman, M. E. (2001). The structure of scientific collaboration networks. *Proceedings of the Aational Academy of Sciences, 98*(2), pp. 404-409.

Niknia, M., & Mirtaheri, S.L. (2015). Mapping a decade of linked data progress through co-word analysis. *Webology, 12*(2) pp. 1–14.

Opsahl, T. (2013). Tradic closure in two-mode networks: Redefining the global and local clustering coefficient. *Social Networks, 35*, pp. 1–20 doi:10.1016/j.socnet.2011.07.001.

Phrma. (2015). *Biopharmaceutical research & development: The process behind new medicines*. (Research Paper). Washington DC: PhRMA. Retrieved from http://phrma-docs.phrma.org/sites/default/files/pdf/rd_brochure_022307.pdf

Pritchard, A. (1969). Statistical bibliography or bibliometrics? *Journal of Documentation, 25*(4), pp. 348-349.

USAID. (2015). *Hilirisasi riset untuk kesejahteraan masyarakat*. (Report) .Retrieved on 2 January 2018 from http://www.prestasi-iief.org/index.php/id/feature/104-hilirisasi-riset-untuk-kesejahteraan-masyarakat

Rebholz-Schuhman, D., Cameron, G., Clark, D., van Mulligen, E., Coatrieux, J.-L., Del Hoyo Barbolla, E., ... van der Lei, J. (2007). SYMBIOmatics: Synergies in medical informatics and bioinformatics-exploring current scientific literature for emerging topics. *BMC Bioinformatics, 8* (Suppl 1), S18.

Silalahi, V. M. M., Hardiyati, R., Nadhiroh, I. M., Handayani, T., Rahmaida, R., & Amelia, M. (2018a). A Framework for text mining in scientometric study: A case study in biomedicine publications. *Journal of Physics: Conference Series, 1007*(1), pp. 1–7.

Silalahi, M., Hardiyati, R., Nadhiroh, I. M., Handayani, T., Amelia, M., & Rahmaida, R. (2018b). A text classification on the downstreaming potential of biomedicine publications in Indonesia. In *2018 International Conference on Information and Communications Technology (ICOIACT)* (515–519). Massachusetts: IEEE.

Stegmann, J., & Grohmann, G. (2005). *Transitive text mining for information extraction and hypothesis generation.* Paper presented at the 29th Annual Conference of German Classification Society, Magdeburg, March 9-11, 2005. Retrieved July 10, 2008, from http://arxiv.org/abs/cs/0509020.

Stephen, T. (1999). Computer-assisted concept analysis of HCR's first 25 years. *Human Communication Research, 25*(4), pp. 498–513.

Sukara, E. & Tobing, I.S.L. (2008). Industri berbasis keanekaragaman hayati, masa depan Indonesia. *Vis Vitalis, 1*(2), pp. 1–12.

West, D. (2007). Directions in marketing communication research: An analysis of the international journal of advertising. *International Journal of Advertising, 26*(4), pp. 543–554.