BJ O&PM Brazilian Journal of Operations & Production Management

ISSN 2237-8960 (Online)

ABEPRO ASSOCIAÇÃO BRASILEIRA DE ENGENHARIA DE PRODUÇÃO

RESEARCH PAPER

The investigation of an event-based approach to improve commodities supply chain management

Elliot Maître[1,2], Giovana Ramalho Sena[2], Zakaria Chemli[3], Max Chevalier[1], Bernard Dousset[1], Jean-Philippe Gitto[2], Olivier Teste[1]

[1]University of Toulouse, IRIT, Toulouse, France.

[2]Scalian, Toulouse, France.

[3]Scalian, Paris, France.

How to cite: Maître, E., Sena, G.R., Chemli, Z. et al. (2022), "The investigation of an event-based approach to improve commodities supply chain management", *Brazilian Journal of Operations & Production Management*, Vol. 19, No. 2, e20221160. https://doi.org/10.14488/BJOPM.2022.005

ABSTRACT

Goal: Predicting the evolution of commodities price to improve anticipation to supply-chain disruptions is hard. We propose an approach based on an event detection model on data stream to assist experts in such task. The final goal is to report to experts a meaningful description of the most impactful events occurring around the world, to help them in their daily decision-making.

Design / Methodology / Approach: This work results from a cross-fertilization between business management, Information Technology and Computer science. This work relies on an expert analysis and advanced AI engines, including a case study on a specific raw material and a literature review to define the parameters to supervise.

Results: We propose a general architecture based on IT and business synergy. We conduce a general study on the factors influencing raw materials price fluctuations, namely events influencing supply and demand of the commodity. Finally, we present a case study of the events, which historically affected phosphates prices.

Limitations of the investigation: An in-depth knowledge of the domain is needed to analyze and quantify the events impact on the supply chain.

Practical implications: This approach was first designed for assisting raw material purchasers but it can potentially be reproduced to assist other decision-makers.

Originality / Value: We propose a new approach on how to anticipate the implications of external events on supply chain disruption and raw materials price evolution. This method is multidisciplinary, involving expert domain knowledge and state-of-the-art artificial intelligence.

Keywords: Supply-chain; Commodities Buying; Event Detection; Artificial Intelligence; Multidisciplinary Approach.

INTRODUCTION

The commodity market is the basis of many industrial and consumer goods supply chains (SC) around the world, being responsible for moving 30% of the world's goods (Radetzki and

Financial support: ANRT CIFRE Grant N° 2018/1745.
Conflict of interest: The authors have no conflict of interest to declare.
Corresponding author: elliot.maitre@irit.fr
Received: 31 January 2021.
Approved: 18 January 2022.
Editor: Julio Vieira Neto

Wårell, 2016). The market volatility has been particularly addressed by different research efforts because of its intrinsic complexity of products, producers, geopolitical, economics and weather conditions etc. (Huchet-Bourdon, 2011). In this article, we propose an event detection method based on Big Data and Machine Learning techniques, which will assist purchasers and supply chain managers on decision-making.

Considering the tremendous amount of information exchanged online, it is not easily possible to keep track of everything. However, these external events can have direct impact on the supply chain and raw materials price, but are currently under exploited since existing IT solutions are not able to address this problem (Leveling et al. 2014). Fan et al. (2015) provided a novel direction on using Big Data and Machine Learning technologies on supply chain management, highlighted the complexity and need of using advanced analytics for external uncertainty analyses. This way, the number of unpredictable low-frequency and high-impact events, also called "black-swans" by Taleb (2015) can be reduced with these techniques.

In this context, exploiting external data can be crucial to understand events happening around the world. External data refers to public news, social media, weather records, policies of economics, politics, industries, etc. (Fan et al., 2015). Although traditional news carry a lot of information about events happening around the world, social media are also an important source of information. Indeed, several studies on numerous domains show the predictive power of social media, especially Twitter (Imran et al., 2015; Hasan et al., 2018). Events are discussed and people react in real-time to what is happening (Sakaki et al., 2010). Moreover, Twitter can help to predict the stock market (Bollen et al., 2011; Oliveira et al., 2017) and the activity on social network can be correlated with variation of the stock market (Ruiz et al., 2012; Chen et al., 2014). Newsfeeds such as Reuters can be combined with social media analysis for event detection (Keane et al., 2015). Finally, recent events such as the Gamestop stock rise caused by users of the subreddit "r/WallStreetBets" and the variations of stocks caused by Elon Musk's multiple tweets keep demonstrating the growing influence of social media on the stock market.

Our goal is to provide an event detection method that can detect events that might impact supply chains, with emphasis on the commodities and raw materials purchasing. To do so, we propose a multidisciplinary approach, combining event detection on text data stream and an in depth study of events that have an impact on commodities prices. In the rest of this paper, we will present a general study about events that caused past disruptions on raw materials prices and use the results to calibrate the filtering component of our event detection method. Then, we propose an application of this study on a particularly strategic commodity: the phosphates.

LITERATURE REVIEW

Factors influencing the commodity market

Classical microeconomics assumes that disturbs in the supply are caused by a change in the production costs, mostly due to technology advances, taxes, subventions. The consumer's income, price of substitutes and/or complementary products, and the number of buyers influence demand (Pindyck et al., 2013). Jacks et al. (2011) presented empirical evidence from 1700 to present indicating that commodity prices have historically been more volatile than those of manufactured goods have. The same study finds that volatility has not increased over time; on the contrary, globalization and the integration of the world market have led to less volatility compared to situations of economic isolation.

According to Radetzki and Wårell (2016), in the short term, the balance between supply and demand defines the price of commodities. For a matter of definition, raw materials are often considered the output of the primary sector, comprising agriculture (hunting, forestry and fishing), mining (including fossil fuels) and utilities. However, this definition is narrowed

and a broader and more used classification uses the Standard International Trade Classification (SITC), designed by the United Nations, sections and divisions, allowing to find intrinsic characteristics of groups. This division takes into consideration group (A) for "Food in a broad sense" referred to SITC Section 0 + 1 + 22 + 4, group (B) for "Agriculture commodities" Sections 2 – 22 – 27 – 28, group (C) for "Minerals and Metals" Sections 27 + 28 + 67 + 68 and finally, group (D) for "Mineral fuels" referred to section 3.

Despite the presence of different materials in the groups, most of the substitutes for each material is founded in the same group, which is more evident for the fuels' group. Radetzki and Wårell (2016) premise that for agricultural products, (A) and (B), price instability is more often caused by disturbances on the supply side and the cultivation of certain products is highly geographically concentrated. However, price fluctuations for minerals, (C) and (D), are usually on the supply side - notwithstanding strikes and cartels.

Another raising question is the relation between mineral fuels price peaks and resource depletion. Henckens et al. (2016) investigated the relationship between the price trend of mineral raw materials and the availability of its resources for future generations, by analyzing the market mechanisms. He concluded that despite the fluctuation in mineral resource prices, there is no significant correlation with resource depletion, but with the balance dynamics of supply and demand.

Finally, Anani (2019) studied the long-term sustainability for countries relying exclusively on the commodities market. The author defends that the supply-demand balance dictate commodities price, together with the technical progress and business structure. Plus, the study highlights the growing influence of speculation due to the financialization of raw materials market.

The main findings are that although commodity's price are volatiles, a balance between supply and demand usually settles them. Supply and demand are affected by internal and external events that can disturb the system balance and then influence the price. For the first, disturbs are mainly due to a variation on production costs and for the latter, the price of substitutes and/or complementary products, the number of buyers and their income. In addition, some macroeconomic conditions, such as the speculation and energy prices also affects the price.

The concept of event for SC and Text mining

The definition of an event is not standardized across domains. From a supply chain point of view, this problematic has been mostly addressed by a risk management approach, using Complex Event Processing (CEP) in integration with RFID (Radio-Frequency IDentification) devices, barcodes readers and sensors, which aims to detect events within the supply chain (Leveling et al., 2014). External data is also relevant for managing SC uncertainties, and refers to data collected from public news, social media, weather and natural disasters records, policies on economics, politics, etc (Fan et al., 2015). However, these sources are usually carrying information and some of it is contained in text. Hence, text mining techniques are needed to extract value from external data.

From a computer science point of view, event detection is intimately linked with Topic Detection and Tracking (TDT), which is a classical text mining task (Allan, 2012), and can be considered as a subtask of TDT. However, detecting a topic is not necessarily detecting an event. We want to detect events that might disrupt the supply chain and not just understand what the discussions are in the data stream. Hence, we need to detect discussions that are linked with real-world events. Such events can include several types of events, such as earthquakes (Sakaki et al., 2010), floods (Avgerinakis et al., 2017) or locust invasions (Tabar et al., 2021). The definition given by McMinn et al. (2013) is close to what we are looking for: "An event is a significant thing that happens at some specific time and place". They complete this definition with "Something is significant if it may be discussed in the media. For example, you may read a news article or watch a news report about it". We will have to characterize the events, which means by using for example the "5W1H" technique, discussed by Xiang and Wang (2019): "What,

Where, When, Who, Why, How". This is a difficult task, particularly when working with streams extracted from social media where information is fragmentary.

Thus, we will focus on detecting events on data streams, with a special emphasis on detecting events that might cause supply chain disruption and particularly variations in the price of raw materials.

Event detection using text data stream

Social media are widely used for event detection. Twitter is the most used social network for event detection (Hasan et al., 2018). It is the fastest social network to react, with long terms and short-term information (Zubiaga et al., 2018). A common approach to detect events is to work on "bursty features detection" (Fung et al., 2005). Keane et al. (2015) work on the bursts detection in word usage and propose an approach based on the combination of social media and newsfeeds from known newspaper. The issue for our context is that this technique focuses on detecting bursts, which supposes that the event is drawing a lot of attention. However, we will work on a very specialized domain which is not commonly discussed and thus may be drawn in the amount of data. Our goal is closer to one of the task of TDT, cluster detection, which attempts to cluster documents as events using clustering techniques. The clusters of documents match the real events and by extracting information from these clusters, one can answer the "5W1H".

Different approaches for document clusterisation are to be considered: document pivot (Petrović et al., 2010; Mazoyer et al., 2020) and feature pivot (Weng & Lee, 2011; Fedoryszak et al., 2019). The difference between these two approaches is granularity. Feature pivot works at the word level while document pivot is focused on the document level. We will present a document-pivot approach. We consider that the document level approach is more complete because it carries more context by using the whole text and not only keywords. In fact, the articulation of human sentence goes beyond the keywords as it carries a lot of information, and working at the feature level might lose this information.

Objectives of the study

We investigate a framework based on IT and business synergy which will help decision makers by providing insightful information about events impacting commodities stock market in real time. In the literature, there is a lack of an investigation about the data sources to monitor in order to detect the impactful events and about the nature of the events to supervise. We identify the relationship between the historical events and the variations of the commodities stock market to calibrate the event detection model filters, i.e. establish which type of event to monitor (political, weather conditions, geographical localization ...) and which sources to focus on. We apply this analysis to the phosphate as a case study.

METHOD

This research is part of a R&D program conducted by a consulting company focused on operational excellence, with the SC being one of its core competencies. A broader presentation of the IT part of this work is presented in (Maître et al., 2020).-Hence, this work is an effort to propose innovative solutions for different industries. In order to validate the interest of this multidisciplinary approach (computer science, machine learning, supply chain management, economics) and its applicability in the fields of logistics and procurement, we first conducted a series of interviews with experienced supply-chains experts and purchasers.

In total, we consulted five different experts throughout 10 hours in-person/phone interviews. The results of these interviews served as a guidance for this research, by providing the firsts insights for the following questions:

• What type of data sources should we use?

• What type of signals should we monitor?
• How to define a signal?
• Which metrics should we use?

The output of these interviews revealed that information sources are multiple and they are difficult to follow. Events have an impact on raw materials' price but an in-depth study of the dynamics of variation is needed for each commodity. Considering the wideness of the domain and the multitude of raw materials, we concluded that a pilot study had to be performed to assess the feasibility of this approach. The Phosphate was then chosen to perform the rest of the study, since it has been included as a Critical Raw Material by the European Commission, the absence of substitutes, the fact of being a mineral needed for the food system, and of a particular scientific interest.

The method we propose is divided into two major components, the (A) business component and the (B) IT component, which are interdependent and complementary. The general architecture is described in Figure 1. The former, which will also be assimilated as phase 0, is an in-depth study of the events that historically impacted the raw material to be supervised, in our case the phosphate. These results will then be exploited to calibrate the filtering rules of the first phase of our event detection model, the IT component of our method.

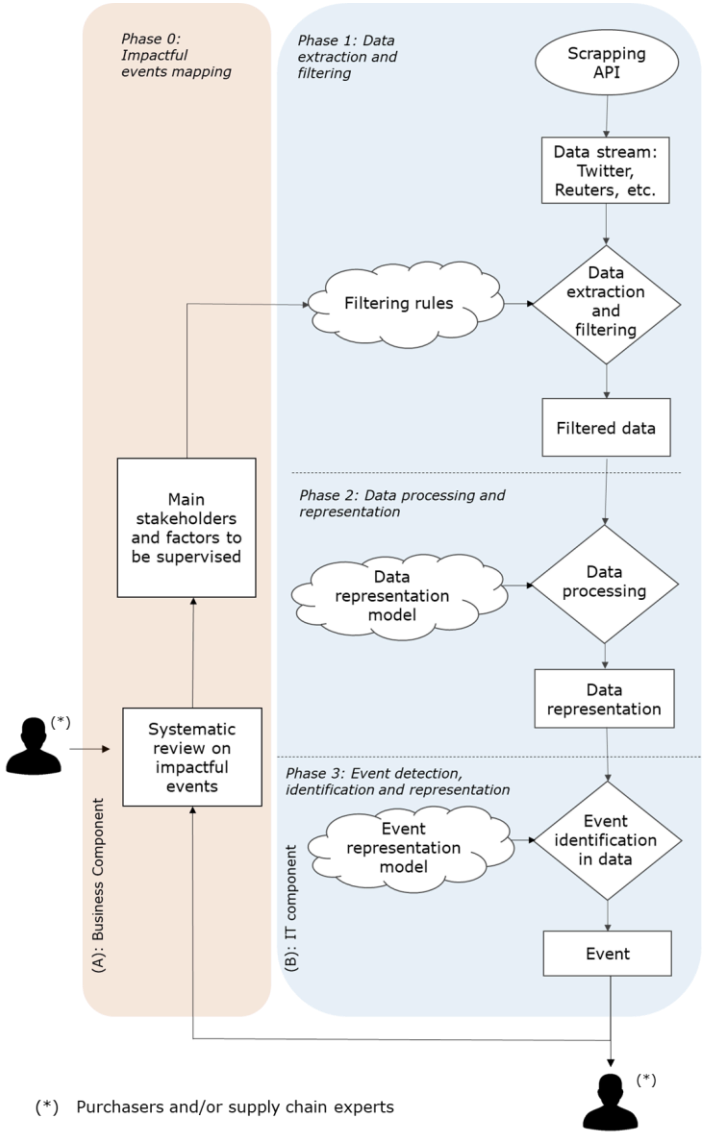

Figure 1 – The general architecture of the event detection pipeline. Source: The authors themselves

This second component is then subdivided in three phases: (1) Data sourcing, extraction and filtering, (2) Data treatment and representation, (3) Pattern recognition and events identification. This second component has one goal, detecting events, which will have a dual utility. These events will be presented to purchasers and supply chain managers so they can analyze them and make the adequate decisions. They will also be taken into account during future occurrences of phase 0, promoting an interactive process to update the important events to supervise and thus the filtering rules, highlighting the duality between the business and IT components.

The rest of this section will be articulated as follow. First, we will describe the business component, also called phase 0. Then, the IT component will be presented, with a description of each of the three phases.

Phase 0 - Proposition for impactful events mapping

To create the data filtering process in Phase 1, it is necessary to identify the events that had historically impacted raw material price. Then, using these results we avoid processing too many information and only consider the relevant ones by using the appropriate filters. As this phase is a building block for the IT component robustness, it is so called Phase 0.

Hence, a literature review was conducted to create a typology of events that has already been identified as triggers and/or reactions for a price change. The following procedure was then executed for the Phosphates case study and could be replicated for other raw materials. In addition, the main actors and their geographical zone were identified to further monitoring, by the analyses of resources and production distribution with data collected from the U.S. Geological Survey, Minerals Yearbook 2018, v.I, Metals and Minerals, as well as an import and export analyses with World Bank data.

To respond to the multidisciplinary needs of this research, Google Scholar was the main search engine used since it indexes a multitude of sources. Most consulted publications were in peer-reviewed journals in the fields of Raw Materials, Resources, Economics and Food Policy. In addition, conference proceedings and books, as well as dissertations, scholars working papers and institutions publications (i.e.: OCDE, World Bank) were selected to enrich the analyses. The following keywords were used: commodity, raw materials, price, supply chain, events, phosphates. Due to the recent emergence of social-media which modified the classical ways of communications, papers published after 2000 were privileged.

Therefore, a non-exhaustive list of the main causes of perturbations in raw materials price were identified and they are presented in a summary table, together with their literature of reference. These causes will be used to setup the filtering rules to detect events that are similar to those which have been identified in the literature. Then, a list of the main actors of the supply chain was elaborated, associated with the role, geographical zone and impact characteristics. These lists will be consistently updated, using the feedback given by the event detection model and the analysis of experts. The quantification of the events in terms of price variation intensity, most probability and their importance is not treated in this paper, being subject for further investigation.

Phase 1. Data sourcing, extraction and filtering

In this phase, we exploit the knowledge gathered from the study of raw materials to implement effective filters for the data sources. Indeed, not all the events are interesting for our context and previously determining which events and sources to focus on is crucial. Multiple data sources can be considered, such as newsfeeds and social media.

Since social networks are not domain-specific, we can consider that not all the discussions through social networks are relevant for SC management improvement. Java et al. (2007) also underlined that not all the tweets are related to events. So, in our context, applied to a specific domain, one of the main challenges is to filter information through the feeds in order to

identify the most relevant information. Thus, it is necessary to develop filtering method in order to reduce the spam and the unnecessary information within the feeds.

Phase 2. Data processing and representation

Once the incoming stream has been filtered, we need to clean the data and extract value from it. There is a lot of information to represent from each tweet such as the text but also the metadata such as date, localization of the tweet, user posting it, etc. (Edouard, 2017). To represent the text content, we use state-of-the-art Transformer based NLP model, BERT (Devlin et al., 2018). This type of model is indeed particularly effective for document representation. We use it to represent the whole text of the document, to embed as much context as possible. The model outputs an embedded vector for each document. Embedded vectors generated for document that have a similar meaning are close. The content of these vectors can be augmented by analyzing the metadata of the document.

Phase 3. Event detection, identification and representation

Thanks to the embedded vector representing the content of each document, we can group documents that are the most similar in different clusters using clustering algorithms (Blondel et al., 2008). Each cluster will represent an event, or a part of an event. Once the clusters are obtained, a representation of every cluster is generated, using the aggregation of the textual content of the cluster (Klas & Fuhr, 2000). Then, this content is summarized, in order to extract the important topics of the cluster to present it to experts. To do so, we use BERT Summarizer (Liu & Lapata, 2019), which is variation of the BERT model presented previously. Finally, to track the evolution of events during time, clusters representations will be successively grouped (Fedoryszak et al., 2019). Once the events are successfully identified, they are proposed to experts so they can analyze them and incorporate them in their decision-making process. These events will also loop back to phase 0, so they can be considered for potential update of the filtering rules. The Technical considerations of the phase 2 and phase 3 will not be developed in this paper. This work is currently being conducted and will be described in the future.

Thus, we proposed a multidisciplinary approach in which the component of each domain complement the other. We will now present the results for the study conducted for phosphates.

RESULTS

As mentioned previously, in the context of this study, phosphate was chosen as the raw material for the development of the pilot case of the detection of weak signals on social networks. For terminological clarification purposes, the difference between phosphorus (P in the general sense and P_4 with reference to commodity) and phosphates result from their chemical composition. They represent the only chemical element and its commodity, respectively, whereas phosphates are compounds when phosphorus is bound to oxygen and other mineral elements (De Ridder et al., 2012). To clarify, phosphate rock in the form of P_2O_5 (i.e. phosphorus pentoxide) is a practical and standard measure of the phosphorus content of any product (Mew et al., 2018) and therefore will be privileged for the rest of the study.

Beyond its essential character for life, it is non-substitutable and exhaustible. Together with nitrogen (N) and potassium (K), they form the bricks of modern fertilizers (NPK), making them one of the key raw materials for food security. In fact, phosphate is mainly used in the production of fertilizers, in animal feed and in food additives and a small fraction in industrial processes. As stated earlier, the phosphate supply chain begins in the mining of phosphate rock. Then, treatments are applied according to the desired use (Blengini et al., 2020). The simplified supply chain of rock phosphate and element phosphorus is illustrated in the Figure 2 below. Likewise, the main by-products of each step of the process, as well as the distribution of the main uses in the European

Union are presented. The main usages for the phosphate rock is to produce fertilizers for agriculture and as an input for animal feed. Despite its non-substitutable character in agricultural applications, the same statement it is not applicable in industrial uses (Heckenmüller et al., 2014). Therefore, this type of application will not be considered for the rest of the study, neither the recycling process, making the phosphates food supply the subject of this research.

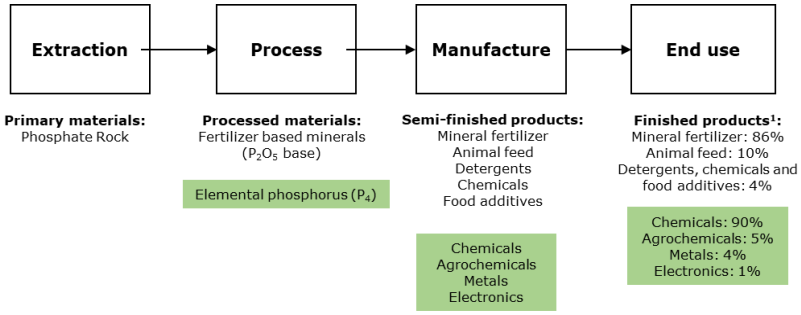

Figure 2. The simplified phosphate supply chain. Adapted from the European Commission 2020.

Thus, the phosphates food supply chain comprises the sectors and processes related to mining, phosphate rock/phosphorus processing and trade, fertilizers production and trade, agriculture application of fertilizers in crop and pastures, food production, processing and distribution and the final consumption. Hence, there is an industry's trend towards vertical integration was highlighted in Van Kauwenbergh (2010).

In the literature review to identify factors affecting raw materials prices, the main findings were that price variations were frequently caused by a perturbation in the supply demand balance. This hypothesis was also used to explain major historical price changes for Phosphates (Mensah, 2003; Heckenmüller et al., 2014; Cordell et al., 2015; Mew et al., 2018). Given the goal of identifying the events impacting the price of phosphate, it is necessary to understand the dynamics of supply and demand of the market. In this way, the world reserves and the main phosphate rock producing countries have been identified and they are presented in the Tables 1 and 2 below. The data presented in the tables below are until 2018 because it was the latest publication until this article was written.

Table 1. Phosphate rock world reserves. Source: Jasinski (2018)

Country	Estimated Reserves (tonnes)	Participation %
Morocco and Western Sahara	50 000 000	71,8%
China	3 200 000	4,6%
Algeria	2 200 000	3,2%
Syria	1 800 000	2,6%
Russia, Peru, Kazakhstan, Tunisia, Uzbekistan, Israel, Senegal, India, Mexico and Togo*	1 713 000	2,5%
Brazil	1 700 000	2,4%
South Africa	1 500 000	2,2%
Saudi Arabia	1 400 000	2,0%
Egypt	1 300 000	1,9%
Australia	1 100 000	1,6%
United States	1 000 000	1,4%
Finland	1 000 000	1,4%
Jordan	1 000 000	1,4%
Other countries	770 000	1,1%
World Total (rounded)	70 000 000	
** Between 0,1% - 1,0% each*		

Table 2. World Production. Source: Jasinski (2018)

Country	Average (2014-2018) in thousand metric tonnes	Participation %
China	38 960	50,6%
Morocco	9 074	11,8%
United States	7 486	9,7%
Russia	4 766	6,2%
Jordan	2 572	3,3%
Brazil	1 996	2,6%
Saudi Arabia	1 571	2,0%
Egypt	1 397	1,8%
Peru	1 123	1,5%
Israel	1 116	1,5%
Tunisia	1 090	1,4%
Vietnam	892	1,2%
Australia	758	1,0%
South Africa, Mexico, Senegal, India, Algeria, Finland, Togo, Turkey, Kasakhstan, Syria, Uzbekistan, Iran, Nauru*	4 039	5,3%
Other countries	86	0,1%
World Total	76926,05	100,0%
** Between 0,1% - 1,0% each*		

It is in Morocco where most of the world's phosphate rock reserves are found (71.8%) and China takes the place of the world's largest producer (50.7%). There is a high concentration of the market, as five countries (China, Morocco, USA, Russia and Jordan) produced around 80% of the phosphates in the world. This highlights a particular dynamic of the world market, which corresponds to the import and export balance of countries, their characterization as producing, exporting or exclusively importing countries. Therefore, an analysis of the trade balance between the countries was carried out on the World Bank data for the main phosphate products including Phosphorus (P_4), Phosphate rock (P_2O_5). The results are shown in Figures 3 and 4 below.

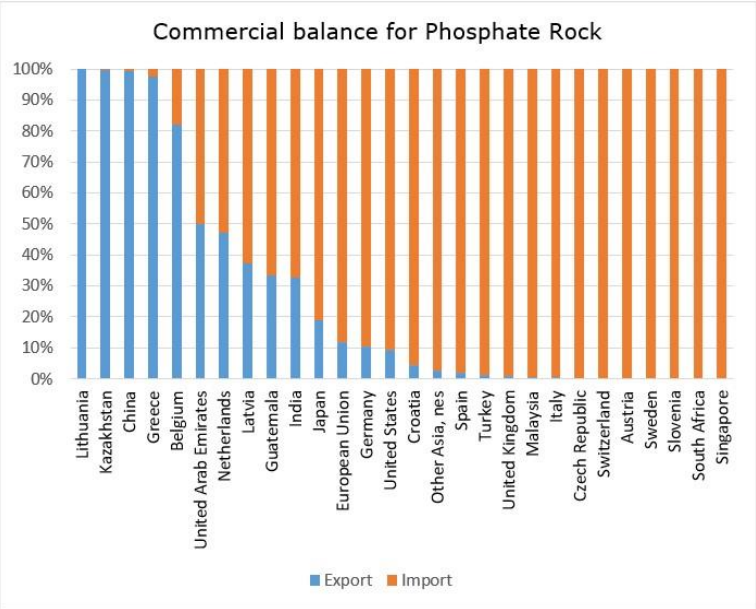

Figure 3. Phosphate rock commercial balance. Source: World Bank Data 2019. Available at: https://wits.worldbank.org/. Accessed on 14/12/2020

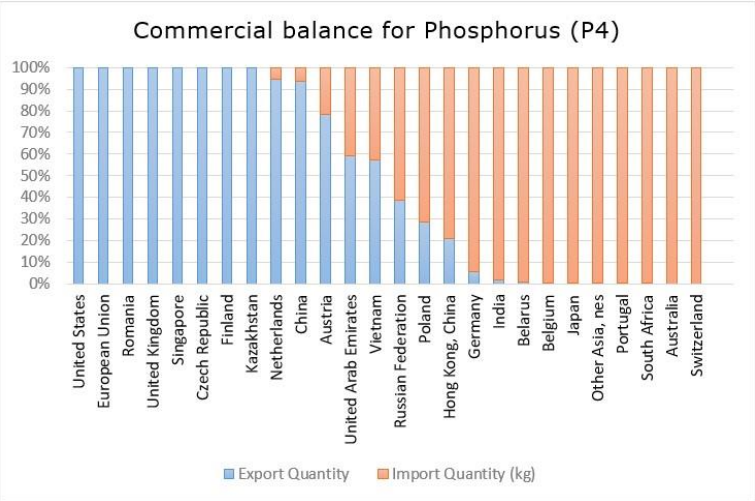

Figure 4. Phosphorus commercial balance. Source: World Bank Data. Available at: https://wits.worldbank.org/. Accessed on 14/12/2020

They confirm the market concentration and dependence of many countries on import, of which about 91% depend completely on the import of phosphate rock and/or phosphorus to meet their internal demand. Since Phosphorus (P_4) is obtained after processing, the main exporting countries are not the same as for the Phosphate Rock. This highlights the dependence of a large amount of the globe that relies on few producers and the concerns on phosphorus availability for food security. (Cordell et al., 2015) On the one hand, monitor events that could disturb production in the producing countries is then crucial for importers to anticipate their response to eventual supply disruptions. On the other hand, if there is a non-expected demand peak, producers should be able to anticipate and invest in production expansion, since the time gap between the investment decision and the operational plant can be of 3 to 5 years. (Weber et al., 2014)

From there, the identification of geographic areas to be supervised as well as the major players in the global market was initiated, in view of the supply chain previously presented. As mentioned before, these actors represent the majority of transactions in the field for the food supply chain, and can contribute in different ways for market balance. They are presented in the form of a simplified stakeholder matrix in the table at in the Annex A, which will serve as a basis for the construction of weak signal filters, with the aim of defining the types of documents that are published and the types of events that we want to search according to the player supervised.

An initial sample of 135 main players was identified and divided by sector, role, geographical zone and if their side in the economic balance (demand, supply or context) for the phosphates use case. For companies from the mining sector and in the production of fertilizers, as they are upstream in the value chain, the mine location was used; in the downstream, for the agribusiness companies, their headquarters' location was considered. Transversal actors from the supply chain, such as business associations, non-governmental organizations, institutions, the academy and specialized international research groups. They were considered as context factors, providing information for the macroeconomics condition of the market.

As Van Kauwenbergh (2010) highlighted the similarities in the price behavior between the fertilizers and the phosphates, as well as the industry trend to verticalisation, their intrinsic relationship should be exploited. Hence, members from the International Fertilizers Association (IFA) were also identified for further monitoring. They represent more than 400 institutions from 72 different countries.

Finally, once the main players and the geographic areas have been identified, the chronological analysis of the phosphate price was carried out through a review of the state of

the art on the historical peaks of phosphorus. To create a typology of events that have impacted the price and could impact it again in the future, they are categorized as affecting supply, demand and / or the entire system, related to the economic context. The events impact period (short or long-term variations) is also highlighted. Factors identified as risks to food security by Cordell et al. (2015) such as political instability were also considered with regard to its influence on the availability of supply. The Table 3 below presents the outcomes of this analysis, inspired by the work of Rezitis and Sassi (2013) for commodity food prices.

Table 3. Events that influence Phosphates price. Adapted from Rezitis and Sassi (201

Potentiel triggers	Economic rationale	Time frame	Reference
	Supply factors		
Exportation politics	Somer of the exporting countries introduce restrictive trade policies aimed at isolating their economies and controlling the market. Example: China in 2008	Short-term	(Cordell and White, 2015)
Low investment in R&D and slower infrastructure response	Limitation of production capacities: the gap between the investment decision and the actual production is about 3 to 5 years	Long-term	(Weber et al., 2014; Heckenmüller et al., 2014)
Political instability in producing countries	Examples: Tunisia during the Arab Spring and in Syria today due to the ongoing civil war (De Ridder et al., 2012)	Short-term	(Heckenmüller et al., 2014; Cordell et al., 2015; Cordell and White, 2015)
Higher production costs	Energy prices, labor, costs, costs of chemical reactants (e.g. water, price of sulfur and sulfuric acid)	Medium-term	(Mew et al., 2018)
	Demand factors		
Weather an crop quality	Agriculture production shortfalls due to adverse weather conditions lead to lower global food supply and higher fertilizers demand	Short-term	(Mensah, 2003; International Fertilizer Industry Association, 2011)
Emerging Economies	Increased income in the BRICs and more urban population is changing food habits pushing demand for higher-value products	Long-term	(Cordell et al., 2015; Cordell and White, 2015)
Biofuels productions and politics	Increased demand for crops used as inputs in the production of biofuels	Long-term	(Cordell et al., 2015)
Food commodity prices	Encourage farmers to increase their crop yields by applying more fertilizers, including phosphate fertilizers	Short-term	(International Fertilizer Industry Association, 2011; Huchet-Bourdon, 2011)
Importation politics	Protectices, import tax to promote the domestic economy	Short-term	(Cordell et al., 2015)
	Market and macroeconomic conditions		
USD change rate	The purchasing power of countries varies according to its currency	Short-term	(Huchet-Bourdon, 2011; Cordell et al., 2015)
Speculation and finacialization	Reference price on the stock markets, which influences production and consumption decisions	Short-term	(Cordell and White, 2015; Anani, 2019)
Energy and fertilizers price	Inputs for agriculture: increase in production costs; increased production of biofuels	Medium-term	(Heckenmüller et al., 2014; International Fertilizer Industry Association, 2011; Huchet-Bourdon, 2011)
Food commodity prices	The price of phosphate rock follows food and agricultural prices very closely, although they lag slightly behind (a month or two)	Short-term	(Van Kauwenbergh, 2010; Heckenmüller et al., 2014)

On the one hand, phosphates supply is mainly influenced by factors related to production disruptions due to higher production costs (ex: energy and sulfuric acid). In addition, the highly production concentration in political instable countries put in evidence the need to monitor political events, as well as exportation policies. On the other hand, short-term triggers for a demand peak could be related to production shortfalls due to extreme weather conditions. Biofuels politics is also an interesting point to be looked to, since it can burst the fertilizer demand to grow biofuel crops. (Cordell et al., 2015).

Events on market and macroeconomics conditions are those who characterize the economic period and context. Main findings in the literature were related to the USD exchange rate, since all contracts are negotiated in this currency; there is a growing influence of speculation and the financialization of the commodities market in the literature. Anani (2019) defends that this movement started by the Commodity Future Modernization Act (CFMA), when stock investors from the USA were allowed to place their assets in raw materials. The economic function of speculation is to mitigate price variations due to changes in supply or demand, this mechanism influences the commodities and also the phosphates market.

Different studies put in evidence the risks dynamics of the phosphates multi-stakeholder supply chain. Cordell et al. (2015) identified different risks throughout the food supply chain and its link to phosphates availability, how they are transferable to other layers. It reveals an interdependence of the steps and actors of the supply chain, which creates a systems feedback effect that was also featured in Wellmer and Dalheimer (2012) and Heckenmüller et al. (2014) and should be taken into consideration for the event detection. Thus, this is also evidentiated by the similar behavior and trend between food, fertilizers and phosphates prices previously mentioned.

Major long-term trends, such as changes in consumers' income and therefore in their diets should be less exploited for further analysis since short-term events are privileged. The risk of resource depletion will not be considerate either, since it has been proved as having no real connection with price variations (Heckenmüller et al., 2014) and due to the differences between existing current estimations and technological developments for optimizing industrial processes. The growing mention of the influence of stock market financial speculation on commodity prices will be further investigated for the remainder of the study.

In sum, the main events type to be supervised will be related to the countries commercial politics (ex: export or import taxes, biofuels), political events (war, demonstration, riots), weather conditions in major crop areas, the food commodities market, as well as the sulfuric acid market; major financial indicators, such as the USD rate and the financial markets. These events whether affect the supply or the demand of phosphates and could then; influence its prices. They were identified after a literature review on phosphates, but this approach could be extended to other commodities.

Most studies on raw materials analyze historical prices within the framework of an economic analysis of raw materials' volatility (Huchet-Bourdon, 2011), the challenges of sustainable development (Cordell et al., 2015; Anani, 2019), as well as the issue of price in relation to the resources' depletion (Henckens et al., 2016). Events have been identified as being the causes of price variations through statistical correlation and causality tests, notably on the effects of the price of fertilizers, crude oil, agricultural commodities, and the USD exchange rate. (Huchet-bourdon, 2011) Nevertheless, real-time event detection from social networks and news media for commodity price prediction has not yet been the subject of any referenced study.

Open points in the literature include the impact of biofuels (ie: bioethanol and biodiesel) and speculation. For the first, despite the difficulty of comparison between existing studies due to the different methodologies adopted, most authors agree on the fact that the expansion of the consumption of biofuels has an upward impact on the price of food (Rezitis and Sassi, 2013). For the latter, there is an inherent complexity as pointed by Palazzi et al. (2020) that even sophisticated models are not capable of capturing causal effects to speculation and the markets dynamics.

For future works, it should be privileged to search for these events within the players previously identified, considering their main communication channels, but also the environment they are in. This could be particularly interesting in producing countries for monitoring eventual political events that could disturb supply, as mentioned by Cordell et al. (2015).

For those countries relying exclusively on importation, it could be interesting to monitor their local weather conditions and consequent crop quality to be aware of eventually demand peaks and anticipate a supply chain responsiveness. For industrials from all sectors, additional events could be detected throughout the actors twitter account (when applicable) and monitoring the main institutional publications on the subject. For example, following the latest developments in infrastructure that could increase capacity or improve recycling methods.

Further research efforts will be dedicated to the quantification of these events' impact on the phosphates prices, to provide a trend to the event and facilitate decision-making. Specific investigation on the speculation should also be addressed, since its inherent complexity and causality effects in multiple markets. Recent technology developments such as cryptocurrencies could also be a subject for further research as it increases the possibilities of transactions.

CONCLUSION

This paper investigates an approach that integrates event detection on social networks in real time as an innovative solution to help supply chain decision makers taking informed decisions about future commodities stock market variations. We propose a general architecture based on IT and business synergy, combining NLP and Machine Learning techniques to purchasing expertise. Then, a general study on the factors influencing raw materials price fluctuations was carried out through a literature review. We found that events affecting the commodities stocks are those linked with demand and supply of the commodities and macroeconomic conditions.

Finally, a case study on phosphate was conducted using the previous results. . We investigated institutional sources such as world bank, USGS, IFA to identify the main typology of impactful events (political, weather condition, food supply chain, currencies variations and petroleum prices) and determinate private companies and geographical locations of interest.

Future works will be consecrated to apply the results of this study in the event detection engine development. It consists on building the event search engine on social media data stream to detect events in real-time, which will potentially impact the price of phosphate. For this, the tables of actors to be supervised, as well as the main categories of events affecting the supply or demand will be used. At the same time, a study on the effects of the speculative derivative market on the price of commodities will be carried out to better understand this mechanism and its relationship to weak signals in social networks. The replicability and transposition of the process for identifying events impacting not only other raw materials, but also other components of any supply chain will be studied in more depth.

REFERENCES

Allan, J. (2012), *Topic Detection and Tracking: Event-based Information Organization*, Vol. 12, Springer Science & Business Media, New York.

Anani, E. (2019), *Fluctuations des Cours des Matières Premières: Enjeux de Soutenabilité-Application à l'Afrique de l'Ouest,* Ph.D. dissertation, Université Paris-Saclay (ComUE).

Avgerinakis, K., Moumtzidou, A., Andreadis, S. et al. (2017), "Visual and textual analysis of social media and satellite images for flood detection", in *Multimedia Satellite Task MediaEval 2017*, Dublin, MediaEval.

Blengini, G., Latunussa, C., Eynard, U. et al. (2020), *Study on the EU's list of Critical Raw Materials*, Publications Office of the European Union, Luxembourg.

Blondel, V.D., Guillaume, J.-L., Lambiotte, R. et al. (2008), "Fast unfolding of communities in large networks", *Journal of Statistical Mechanics*, Vol. 2008, pp. P10008.

Bollen, J., Mao, H. and Zeng, X. (2011), "Twitter mood predicts the stock market", *Journal of Computational Science*, Vol. 2, pp. 1-8.

Chen, H., De, P., Hu, Y.J. et al. (2014), "Wisdom of crowds: the value of stock opinions transmitted through social media", *Review of Financial Studies*, Vol. 27, pp. 1367-403.

Cordell, D. and White, S. (2015), "Tracking phosphorus security: indicators of phosphorus vulnerability in the global food system", *Food Security*, Vol. 7, pp. 337-50.

Cordell, D., Turner, A. and Chong, J. (2015), "The hidden cost of phosphate fertilizers: mapping multi-stakeholder supply chain risks and impacts from mine to fork", *Global Change, Peace & Security*, Vol. 27, pp. 323-43.

De Ridder, M., De Jong, S., Polchar, J. et al. (2012), *Risks and Opportunities in the Global Phosphate Rock Market: Robust Strategies in Times of Uncertainty*, Centre for Strategic Studies, Hague.

Devlin, J., Chang, M.-W., Lee, K. et al. (2018), "Bert: Pre-training of deep bidirectional transformers for language understanding", *arXiv preprint,* arXiv:1810.04805.

Edouard, A. (2017), *Event detection and analysis on short text messages,* Ph.D. dissertation, Université Côte D'Azur.

Fan, Y., Heilig, L. and Voß, S. (2015), "Supply chain risk management in the era of big data", in Marcus, A. (Ed.), *Design, User Experience, and Usability: Design Discourse*, Springer, Cham, pp. 283–294.

Fedoryszak, M., Frederick, B., Rajaram, V. et al. (2019), "Real-time event detection on social data streams", in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, Association for Computing Machinery, New York, pp. 2774–2782.

Fung, G.P., Yu, J.X., Yu, P.S. et al. (2005). "Parameter free bursty events detection in text streams", in *Proceedings of the 31st international conference on Very large data bases*, ACM, New York, pp. 181–192.

Hasan, M., Orgun, M.A. and Schwitter, R. (2018), "A survey on real-time event detection from the Twitter data stream", *Journal of Information Science*, Vol. 44, pp. 443-63.

Heckenmüller, M., Narita, D. and Klepper, G. (2014), "Global availability of phosphorus and its implications for global food supply: an economic overview", *Kiel Working Paper*, No. 1897, pp. 1-27.

Henckens, M.L., Van Ierland, E.C., Driessen, P.P. et al. (2016), "Mineral resources: geological scarcity, market price trends, and future generations", *Resources Policy*, Vol. 49, pp. 102-11.

Huchet-Bourdon, M. (2011), *Agricultural Commodity Price Volatility: An Overview*, OECD Publishing, Paris. (OECD Food, Agriculture and Fisheries Papers, No. 52).

Imran, M., Castillo, C., Diaz, F. et al. (2015), "Processing social media messages in mass emergency: a survey", *ACM Computing Surveys*, Vol. 47, pp. 1-38. [CSUR]

International Fertilizer Industry Association – IFA. (2011), *Food Prices and Fertilizer Markets: Factors Influencing the Variations in Fertilizer Market Conditions*, IFA, Paris.

Jacks, D.S., O'rourke, K.H. and Williamson, J.G. (2011), "Commodity price volatility and world market integration since 1700", *The Review of Economics and Statistics*, Vol. 93, pp. 800-13.

Jasinski, S.M. (2018), *Phosphate Rock Statistical Information: Minerals Yearbook*, USGS, USA.

Java, A., Song, X., Finin, T. et al. (2007), "Why we twitter: understanding microblogging usage and communities", in *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis*, Association for Computing Machinery, New York, pp. 56–65.

Keane, N., Yee, C. and Zhou, L. (2015), "Using topic modeling and similarity thresholds to detect events", in *Proceedings of the The 3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, Association for Computational Linguistics, Denver, pp. 34-42.

Klas, C., & Fuhr, N. (2000), A new effective approach for categorizing web documents. https://www.researchgate.net/publication/2551992_A_new_Effective_Approach_for_Categorizing_Web_Documents

Leveling, J., Edelbrock, M. and Otto, B. (2014), "Big data analytics for supply chain management", in *2014 IEEE International Conference on Industrial Engineering and Engineering Management*, IEEE, New Jersey, pp. 918-922.

Liu, Y. and Lapata, M. (2019), "Text summarization with pretrained encoders", *arXiv preprint,* arXiv:1908.08345.

Maître, E., Chemli, Z., Chevalier, M., Dousset, B., Gitto, J. P., and Teste, O. (2020). "Event detection and time series alignment to improve stock market forecasting", In *Joint conference of the information retrieval communities in europe (CIRCLE 2020)*, Samatan, France, pp. 1-5.

Mazoyer, B., Cagé, J., Hervé, N. et al. (2020), "A french corpus for event detection on twitter", in *Proceedings of the 12th Language Resources and Evaluation Conference*, European Language Resources Association, Marseille, pp. 6220–6227.

McMinn, A.J., Moshfeghi, Y. and Jose, J.M. (2013), "Building a large-scale corpus for evaluating event detection on twitter", in *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, Association for Computing Machinery, New York, pp. 409–418.

Mensah, A.A. (2003), "Dynamique et comportements stratégiques sur le marché international du phosphate", *Mondes en Developpement*, Vol. 2, No. 122, pp. 37-56.

Mew, M.C., Steiner, G. and Geissler, B. (2018), "Phosphorus supply chain—scientific, technical, and economic foundations: a transdisciplinary orientation", *Sustainability*, Vol. 10, pp. 1087.

Oliveira, N., Cortez, P. and Areal, N. (2017), "The impact of microblogging data for stock market prediction: using Twitter to predict returns, volatility, trading volume and survey sentiment indices", *Expert Systems with Applications*, Vol. 73, pp. 125-44.

Palazzi, R.B., Pinto, A.C., Klotzle, M.C. et al. (2020), "Can we still blame index funds for the price movements in the agricultural commodities market", *International Review of Economics & Finance*, Vol. 65, pp. 84-93.

Petrović, S., Osborne, M. and Lavrenko, V. (2010). Streaming first story detection with application to Twitter, in *Human language technologies: The 2010 annual conference of the north american chapter of the association for computational linguistics*, Association for Computational Linguistics, Stroudsburg, pp. 181–189.

Pindyck, R.S., Rubinfeld, D.L. and Rabasco, E. (2013). *Microeconomia*, Pearson, Italia.

Radetzki, M. and Wårell, L. (2016), "The geography of commodity production and trade", In Radetzki, M. and Wårell, L. (Ed.), *A Handbook of Primary Commodities in the Global Economy*, 2nd ed., Cambridge University Press, Cambridge, pp. 28–55.

Rezitis, A. N. and Sassi, M. (2013), "Commodity food prices: review and empirics". *Economics Research International*, Vol. 2013, pp. 694507.

Ruiz, E.J., Hristidis, V., Castillo, C. et al. (2012), "Correlating financial time series with micro-blogging activity", in *Proceedings of the fifth ACM international conference on Web search and data mining*, Association for Computing Machinery, New York, pp. 513–522.

Sakaki, T., Okazaki, M. and Matsuo, Y. (2010), "Earthquake shakes twitter users: real-time event detection by social sensors", in *Proceedings of the 19th international conference on World wide web*, Association for Computing Machinery, New York, pp. 851–860.

Tabar, M., Gluck, J., Goyal, A. et al. (2021), "A PLAN for Tackling the Locust Crisis in East Africa: Harnessing Spatiotemporal Deep Models for Locust Movement Forecasting", in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, Association for Computing Machinery, New York, pp. 3595–3604.

Taleb, N.N. (2015), *The Black Swan: The Impact of the Highly Improbable*, Instaread, San Francisco.

Van Kauwenbergh, S.J. (2010), *World Phosphate Rock Reserves and Resources,* IFDC, Muscle Shoals.

Weber, O., Delince, J., Duan, Y., et al. (2014), "Trade and finance as cross-cutting issues in the global phosphate and fertilizer market", in Scholz, R.W. (Ed.), *Sustainable Phosphorus Management: A Global Transdisciplinary Roadmap*, Dordrecht, Springer, pp. 275-294.

Wellmer, F.-W. and Dalheimer, M. (2012), "The feedback control cycle as regulator of past and future mineral supply", *Mineralium Deposita*, Vol. 47, pp. 713-29.

Weng, J. and Lee, B. (2011), "Event detection in twitter", *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 5, No. 1, pp. 401-8.

Xiang, W. and Wang, B. (2019), "A survey of event extraction from text", *IEEE Access: Practical Innovations, Open Solutions*, Vol. 7, pp. 173111-37.

Zubiaga, A., Aker, A., Bontcheva, K., Liakata, M., & Procter, R. (2018). Detection and resolution of rumours in social media: A survey. *ACM Computing Surveys (CSUR)*, Vol. 51, No. 2, pp. 1-36.

Author contributions: All the authors contributed equally to this paper.

ANNEX A

Impact side	Role	Sector	Player	Geographical zone
Context	Académie et Recherche	Agriculture	INRA	France
		Phosphore	Global Phosphorus Research Initiative	International
			Global TraPs	International
	Institution, Association	Bank	World Bank	International
		Fertilizers	IFA	International
		Bank	IFM	International
		Raw materials	ERMA - European Raw Materials Alliance	Europe
		Ressources research	USGS	USA
Demand	Industry	Agro business	Mars	USA
			Agropur Cooperative	Canada
			Ajinomoto	Japan
			Anheuser-Busch InBev	USA
			Archer Daniels Midland Company	USA
			Arla Foods	Europe
			Asahi Group	Japan
			Associated British Foods	UK
			Bacardi	USA
			Barry Callebaut	Switzerland
			Boparan Holdings	UK
			Brf Brasil Foods	Brazil
			Bunge	USA
			Campbell Soup Company	USA
			Cargill	USA
			Carlsberg	Dannemark
			China Mengniu Dairy Company	China
			CHS Inc.	USA
			Coca-Cola Bottlers Japan	Japan
			Coca-Cola European Partners	Europe
			Coca-Cola HBC	Europe
			ConAgra Brands	USA
			Constellation Brands	USA
			Dairy Farmers of America	USA
			Danish Crown	Dannemark
			Danone	Europe
			Dean Foods Company	USA
			Diageo	UK
			DMK Deutsches Milchkontor	Europe
			Dole Food Company, Inc.	USA
			E & J Gallo Winery	USA

Impact side	Role	Sector	Player	Geographical zone
			Femsa	USA
			Ferrero	Europe
			Flowers Foods	USA
			Fonterra	New Zealand
			General Mills Inc.	USA
			Grupo Bimbo (Mexico)	USA
			Hangzhou Wahaha Group	China
			Heineken	Europe
			Hormel Foods Corporation	USA
			Ingredion Inc.	USA
			Ito En	Japan
			Itoham Yonekyu	Japan
			J R Simplot	USA
			Jacobs Douwe Egberts	Europe
			JBS	Brazil
			Kellogg Company	USA
			Kerry Group	Europe
			Keurig Dr Pepper	USA
			Kewpie Corporation	Japan
			Kirin Holdings	Japan
			Kraft Heinz Company	USA
			Lactalis	Europe
			Land O' Lakes Inc.	USA
			Lindt & Sprungli	Switzerland
			LVMH	Europe
			Marfrig Group	Brazil
			Maruha Nichiro Corporation	Japan
			McCain Foods Ltd	Canada
			McCormick Corporation	USA
			Meiji Holdings	Japan
			Molson Coors Brewing Company	USA
			Mondelez International	USA
			Morinaga Milk Industry	Japan
			Muller Group	Europe
			Nestle	Switzerland
			NH Foods	Japan
			Nisshin Seifun Group	Japan
			Nissin Foods Group	Japan
			Nissui	Japan
			Oetker Group	Europe
			Olam International	Singapoure
			OSI Group	USA
			Parmalat	Europe

Impact side	Role	Sector	Player	Geographical zone
			PepsiCo Inc.	USA
			Perdue Farms	USA
			Pernod Ricard	Europe
			Post Holdings	USA
			Red Bull	Europe
			Royal FrieslandCampina	Europe
			Sapporo Holdings	Japan
			Saputo	Canada
			Savencia Fromage & Dairy	Europe
			Schreiber Foods	USA
			Smithfield Foods/WH Group	USA
			Sodiaal	Europe
			Sudzucker	Europe
			Suntory	Japan
			ThaiBev	Thailand
			The Coca-Cola Company	USA
			The Hershey Company	USA
			The JM Smucker Company	USA
			Total Produce	Europe
			Treehouse Foods	USA
			Tsingtao Brewery	China
			Tyson Foods	USA
			Unilever	Europe
			Vion	Europe
			Yamazaki Baking	Japan
			Yili Group	China
Supply	Industry	Mining and processing	Apatit	Kola (Russia)
			China Molybdenum Co.	Ltd Chapadão (Goiàs, Brazil)
			EuroChem	Kovdorskiy GOK (Russia)
			Foskor	Phalaborwa (South Africa)
			Mosaic Co.	Four Corners (Florida, USA)
				Hopewell (Florida, USA)
				South Ford Meade (Florida, USA)
				South Pasture (Florida, USA)
				Wingate Creek (Florida, USA)
			Nutrien (merger of Agrium and Potash Corp.)	Dry Valley (Idaho, US)
				Swift Creek (Florida, US)

Impact side	Role	Sector	Player	Geographical zone
			OCP	Benguérir (Morocco)
				Boucraâ (Sahara)
				Khouribga (Morocco)
				Youssoufia (Morocco)
			P4 Production, LLC.	Blackfoot Bridge (Idaho, US)
			Sinochem Yunlong Co., Ltd.	Aurora (North Carolina, US)
			Vale	Bayóvar (Sechura, Peru)
				Catalão (Goiás, Brazil)
			Yara	Siilinjärvi (Finland)
		Phosphorus Production	5-Continent Phosphorus Co. Ltd.	China
			Changzhou Qishuyan Fine Chemical Co. Ltd	China
			Kazphosphate LLC	Kazakhstan
			Taj Pharmaceuticals Ltd.	India
			UPL Europe Ltd.	India
			Viet Hong Chemical and Trading Co. Ltd	Vietnam
			Yunphos (Taixing) Chemical Co., Ltd.	China