

# UNRAVELLING THE BIOECONOMY - A SEMANTIC SIMILARITY ANALYSIS OF SCIENTIFIC PUBLICATIONS

Daniella F. dos Santos e Silva<sup>1</sup>  
daniella.fartes@hotmail.com  
+55(21)993602871

Bárbara Bressan<sup>1</sup>  
bbressan@cgee.org.br  
+55(61)981998982

Emilly Caroline C. Silva<sup>1</sup>  
esilva@cgee.org.br  
+55(61)982087871

João Pedro F. Arbache<sup>1</sup>  
jarbache@cgee.org.br  
+55(61)999813822

Marcelo Poppe<sup>1</sup>  
mpoppe@cgee.org.br  
+55(61)999740111

<sup>1</sup>Center for Strategic Studies and Management (CGEE) - SCS Quadra 9 - torre C – 4<sup>th</sup> floor, room 401 – 405. Ed. Parque da Cidade Corporate. Brasília. Brazil

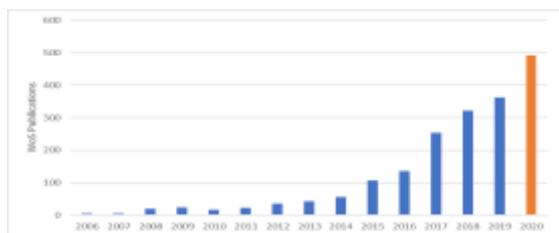
**ABSTRACT:** In the last decade, the term bioeconomy has gained space both in scientific publications and in national strategies and plans as a path to a sustainable, low carbon economy. However, the search for the term in scientific databases still fails to capture its entire field, as it is a new and crosswise theme. Another challenge of studying the bioeconomy at a global level is the influence of each country's regional characteristics on the scope of the bioeconomy. Thus, the objective of this work is to identify which are the main fields of studies considered as part of the bioeconomy, and to analyse how regional characteristics influence the definition of these fields. To achieve this objective, scientific publications of the Web of Science database were studied through semantic similarity methodology using softwares developed by CGEE - InsightNet (iN) and InsightNet Browser (iNB).

**Keywords:** bioeconomy, biobased economy, semantic similarity analysis

## 1 CONTEXT

The term bioeconomy has appeared with increasing frequency both in scientific publications and in national strategies and plans. However, the search in scientific bases for the term still fails to capture its entire area of activity (Fig. 1) because it is a new and quite transversal theme. This article considers the following definition of bioeconomy:

“The bioeconomy comprises all economic activity derived from bioprocesses and bioproducts that contributes to efficient solutions in the use of biological resources - in the face of challenges in food, chemicals, materials, energy production, health, environmental services, and environmental protection - that promote the transition to a new model of sustainable development and social well-being” (CGEE, 2020).



**Figure 1:** Evolution of “bioeconomy” publications.

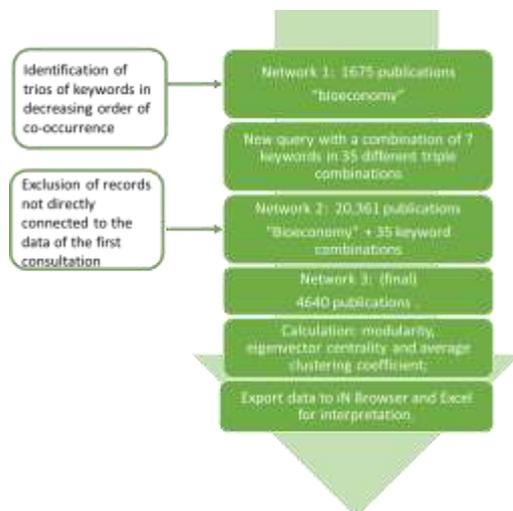
To create strategies and policies for the development of the bioeconomy, it is first necessary to define its scope. Several countries and international institutions have discussed the main areas of the bioeconomy activity, as well as the most relevant types of raw materials, processes and products [2][3][4][5][6]. However, there are still many discrepancies in the use of the term, both in research and policy debates [7]. In this context, the article aims to raise evidence about how the bioeconomy has been discussed in the academic field.

To access a representative sample of publications in bioeconomy, a methodology was built based on the semantic similarity between the publications extracted from the Web of Science (WoS) database, considering the abstracts, titles and the set of keywords provided by the authors and the journals. This methodology will be described below in section 2. In section 3, the results from the generated network will be presented and discussed in terms of thematic clusters, countries that publish the most, and the Brazilian context of publications. Finally, section 4 will present the conclusions of the work.

## 2 SNOWBALL METHODOLOGY

The methodology used to create the network of scientific publications on bioeconomy was based on the

snowball method. Snowball is used to generate search expressions that allow the identification of scientific productions related to a new and comprehensive expression. The steps of the methodology are shown in Fig. 2.



**Figure 2:** Methodological steps

Initially, a network (network 1) was created in the Insight Net<sup>1</sup> program using the 1,675 original publications<sup>2</sup> obtained from the Web of Science (WoS) through the query “bioeconomy”. The 7 keywords that most appeared in co-occurrence trios in network 1 were ranked: biomass, biofuel, bioenergy, biorefinery, sustainability, innovation, circular economy.

Then 35 new queries were made to the WoS using the different combinations in trios of these 7 selected keywords, obtaining a network (network 2) of 20,361 publications. It was then verified which publications from this new network were not directly connected - via semantic similarity analysis - with publications from network 1. Unconnected publications were excluded, forming network 4, with 4,640 articles. From this network, a set of variables was calculated to define the thematic clusters. Next, the program Insight Net Browser<sup>1</sup> and Excel were used to interpret the results. WoS data were collected between August and September 2020.

### 3 RESULTS AND DISCUSSION

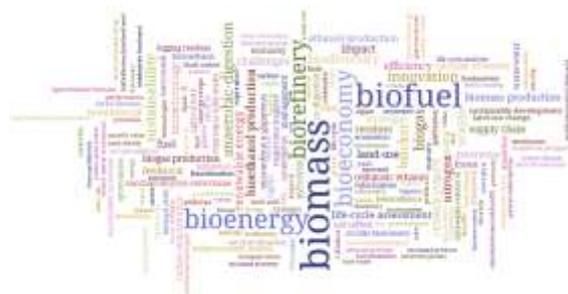
#### 3.1 Network Overview

Fig. 3 shows the keyword cloud (authors and journals) of the final network. The analysis of the most cited keywords points out some relevant indications.

First, there is the word biomass, which is, in fact, the basis of the bioeconomy, which proposes to bring new bio-based raw materials for the production of energy and bioproducts. However, the next two words, biofuels and bioenergy, illustrate energy products, which is clearly just a segment of the bioeconomy. Nevertheless, energy

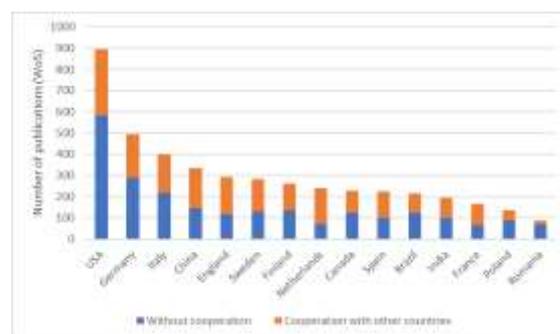
products seem to make up a large part of what has been considered “bioeconomy” by researchers. This result reflects the main use of biomass in a trajectory towards a low carbon economy. However, the fourth and fifth most cited words already point to another movement, which is the identification of new uses of biomass in addition to energy production. The words bioeconomy and biorefinery illustrate this trend.

It is also worth mentioning the presence of the words sustainability and innovation among the 10 most cited, which are concepts strictly linked to the definition of bioeconomy.



**Figure 3:** Network Word Cloud

Fig. 4 shows the 15 countries that publish the most in bioeconomy according to the generated network. The graph also shows the share of publications that was carried out in collaboration with other countries. It is observed that, on average, 48% of publications are made with some level of partnership with other countries. The Netherlands and France stand out here with 70.5% and 60.1% of cooperation, respectively. Brazil was in 11th position with a level of cooperation with other countries of 41.9%.



**Figure 4:** Top 15 countries with the highest number of publications in the network.

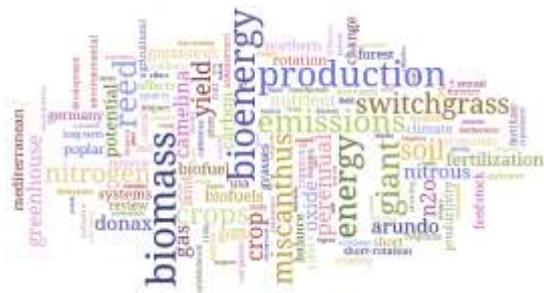
#### 3.2 Thematic Clusters

From the generated network, thematic clusters were selected based on the results of semantic similarity between publications. The 25 largest clusters were considered and within these, we selected the 15 most cohesive. Fig. 5 shows the selected clusters and Table I their themes. To classify and study the clusters, the titles of the publications were analyzed, and the abstracts of the most central publications were read. Next, sections 3.2.1 to 3.2.15 will present a brief description of each cluster, its word cloud, and the 5 countries that publish the most.

<sup>1</sup> CGEE proprietary intelligence tool

<sup>2</sup> We selected the following document types: papers, conference papers, review and early access.





**Figure 8:** Word cloud from cluster 3

**Table IV:** Top 5 countries in cluster 3

Cluster 3 - Grass	
Countries	Total
USA	42
Italy	22
Germany	17
Canada	7
Netherlands	6

### 3.2.4 Cluster 4 – Valorization of waste and coproducts

Cluster 4 deals mainly with the recovery of waste and coproducts. In its central part, there is a concentration of articles focused on the use of sewage sludge for energy production. However, on the periphery of the cluster, there are groups of nodes that deal with the recovery of other residues and co-products, especially glycerol from the production of biodiesel and lignocellulosic residues (Fig. 9).



**Figure 9:** Word cloud from cluster 4

**Table V:** Top 5 countries in cluster 4

Cluster 4 - Valorization of waste and coproducts	
Countries	Total
Poland	12
Italy	8
Canada	7
Sweden	6
Brazil	5

### 3.2.5 Cluster 5 – Algae and lipids

Cluster 5 is large but more dispersed than those previously mentioned. Most publications are focused on topics related to the use of algae for the production of bioenergy and biofuels and also to lipid extraction

processes from algae and microalgae (Fig. 10).



**Figure 10:** Word cloud from cluster 5

**Table VI:** Top 5 countries in cluster 5

Cluster 5 - Algae and lipids	
Countries	Total
USA	15
India	10
China	8
Italy	4
Taiwan	4

### 3.2.6 Cluster 6 – Lignin

The main core of cluster 6 focuses on research on alternative uses of lignin. There is a rich range of topics, ranging from the production of paper, renewable chemicals, and fuels, use in biorefineries, to the physical, chemical, and biological properties of the material. On the periphery of the cluster, we observe related themes such as cellulose processing (Fig. 11).



**Figure 11:** Word cloud from cluster 6

**Table VII:** Top 5 countries in cluster 6

Cluster 6 - Lignin	
Countries	Total
USA	32
Canada	11
China	7
France	5
Brazil	4

### 3.2.7 Cluster 7 – Biochar

Cluster 7 stands out for being very dense and having all its nodes concentrated in the same space within the network, which represents the approach of the same theme: biochar. Biochar is a carbon-rich material used as a soil corrector to improve its quality. The cluster highlights topics such as the different biochar production

methods: pyrolysis, gasification, and hydrothermal conversion; as well as other topics: environmental and economic impacts, regulatory aspects, and risk analysis (Fig. 12).

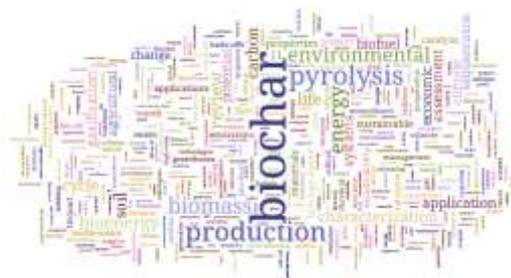


Figure 12: Word cloud from cluster 7

Table VIII: Top 5 countries in cluster 7

Cluster 7 - Biochar	
Countries	Total
USA	19
China	8
Canada	7
[China, USA]	5
South Korea	4

### 3.2.8 Cluster 8 – Palm Oil

Cluster 8 is quite dense and has palm oil as its theme. Studies are mainly divided into reviews, technical and economic analyses, and life cycle analyses. In this cluster, Malaysia stands out as the country with the largest number of publications (Fig. 13).



Figure 13: Word cloud from cluster 8

Table IX: Top 5 countries in cluster 8

Cluster 8 - Palm Oil	
Countries	Total
Malaysia	23
England	23
USA	16
Australia	8
Canada	4

### 3.2.9 Cluster 9 – Straw

This cluster mainly deals with the use of straw from various agricultural products such as sugar cane, corn, and wheat in the generation of energy and the

transformation into bioproducts. In addition to the work on the use of straw, there is also a strong presence of themes aimed at the availability of this biomass and at assessing soil quality after the collection of straw. Also noteworthy is the presence of Brazil as the country that publishes the most on the topic. Brazilian publications are mainly focused on the use of sugarcane straw (Fig. 14).



Figure 14: Word cloud from cluster 9

Table X: Top 5 countries in cluster 9

Cluster 9 - Straw	
Countries	Total
Brazil	9
USA	8
China	7
England	6
Denmark	4

### 3.2.10 Cluster 10 – Microalgae and wastewater treatment

Cluster 10 is very cohesive, focusing on the use of microalgae in the treatment of effluents and subsequent transformation into energy and bioproducts. This cluster is very close to cluster 5 (algae and lipids), but the specificity of the topic in question was enough to generate a new modularity class. India stands out as the country that most publishes on the subject, while Brazil holds the fourth position (Fig. 15).



Figure 15: Word cloud from cluster 10

Table XI: Top 5 countries in cluster 10

Cluster 10 - Microalgae and wastewater treatment	
Countries	Total
India	9
USA	8
Malaysia	7

Brazil	6
China	6

### 3.2.11 Cluster 11 – Hydrogen

This cluster focuses on the production of hydrogen, mainly by biomass gasification, but also by other processes such as anaerobic digestion and fermentation. In addition to the technical publications focused on the hydrogen generation process, many review articles and life cycle analyses were also verified (Fig. 16).



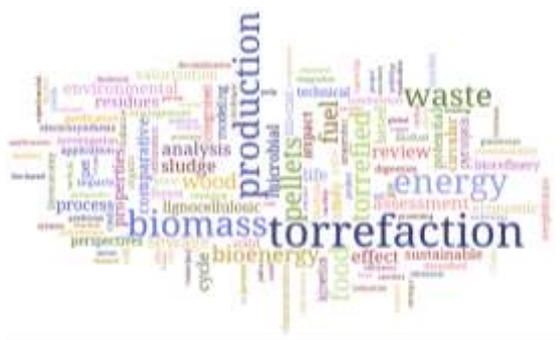
**Figure 16:** Word cloud from cluster 11

**Table XII:** Top 5 countries in cluster 11

Cluster 11 - Hydrogen	
Countries	Total
USA	11
Italy	10
Germany	6
Canada	5
China	5

### 3.2.12 Cluster 12 – Torrefaction

Cluster 12 has as its central theme the biomass torrefaction process. Torrefaction is a thermochemical process that maintains about 75-95% of the energy content. Thus, this cluster deals with this topic, relating it mainly to energy use, the preparation of biomass for gasification processes, and the production of pellets (Fig. 12). The strong participation of the USA and Brazil among the 5 countries that most publish on the subject stands out (Table XIII).



**Figure 17:** Word cloud from cluster 12

**Table XIII:** Top 5 countries in cluster 12

Cluster 12 - Torrefaction	
---------------------------	--

Countries	Total
USA	13
India	5
England	4
Italy	4
Brazil	3

### 3.2.13 Cluster 13 – Miscanthus x Giganteus

The main theme addressed in this cluster is the African and Eurasian plants *miscanthus x giganteus*, which have a large carbon capture capacity. The publications are mainly related to the production of bioenergy and biofuels and the evaluation of the productivity of *miscanthus x giganteus* (Fig. 18).



**Figure 18:** Word cloud from cluster 13

**Table XIV:** Top 5 countries in cluster 13

Cluster 13 - Miscanthus x Giganteus	
Countries	Total
USA	19
China	6
Germany	4
Croatia	3
England	3

### 3.2.14 Cluster 14 – Circular economy

Cluster 14 is strongly characterized by the theme of circular economy and also by “circular bioeconomy”. Within this theme, the publications deal mainly with new business models for bioproducts. It is also worth mentioning a subgroup within the cluster focused on circular economy in the Baltic countries (Fig. 19).



**Figure 19:** Word cloud from cluster 14

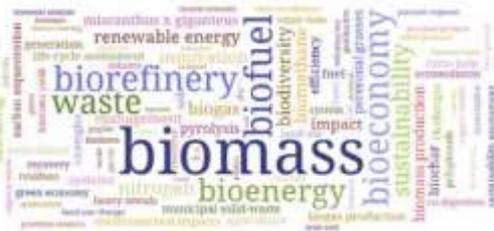




**Figure 22:** Germany word cloud

### 3.3.3 Italy

Fig. 23 shows the keyword cloud for Italian publications. The country produced 399 articles, corresponding to 9% of the network. It is worth mentioning 3 areas where there was a relevant number of publications: use of grass biomass (clusters 2 and 13); production and use of olive oil and its co-products; and processing of sewage and effluents mainly via the anaerobic digestion process.



**Figure 23:** Italy word cloud

### 3.3.4 China

China presented 336 articles on the network (7%), 190 of them (57%) in cooperation with other countries. The country that published the most in partnership with China was the USA, with 27% of the articles in cooperation. The other 2 countries that published the most with China were Canada (5%) and Taiwan (4%). Fig. 24 shows the keyword cloud for Chinese publications. Like Germany, Chinese publications were widely spread across the different clusters of the network, but the relevant presence in the topics of biodigestion of effluents and sewage and production of levulinic acid stands out.



**Figure 24:** China word cloud

### 3.3.5 England

England was responsible for 294 publications on the network (6%), of which 175 (60%) were in cooperation with other countries. The three countries that published the most with England were the USA (5.1%), Italy (4.6%) and

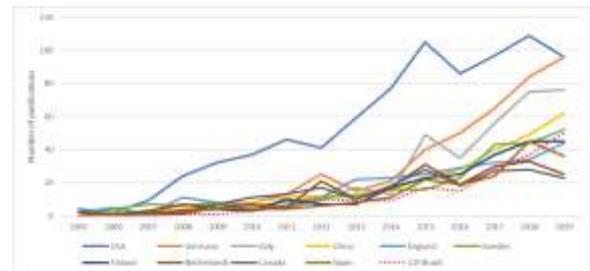
Scotland (4.6%). Fig. 25 shows the keyword cloud for English publications. It is interesting to note in the figure the presence of the words “Brazil” and “sugarcane bagasse”, which reflect a set of publications in partnership with Brazil concerning the use of sugarcane bagasse for the production of second-generation ethanol.



**Figure 25:** England word cloud

### 3.4 Results for Brazil

This section aims to present the results of Brazilian publications on the network. Fig. 26 shows the evolution of the publications of the 10 countries that most published and Brazil (11th place). It is possible to observe the trend of rapid growth that Brazil has had in the theme of the bioeconomy since 2016.



**Figure 26:** Top 11 publishing countries

Fig. 27 shows the keyword cloud for Brazilian publications. From the analysis of the publications, it was possible to observe a focus on the area of production of biofuels and bioenergy, mainly from sugar cane biomass. The predominance of Brazil in cluster 9 reflects this focus. It is worth pointing out, however, that Brazilian publications were widely spread across the generated network.



**Figure 27:** Brazil word cloud

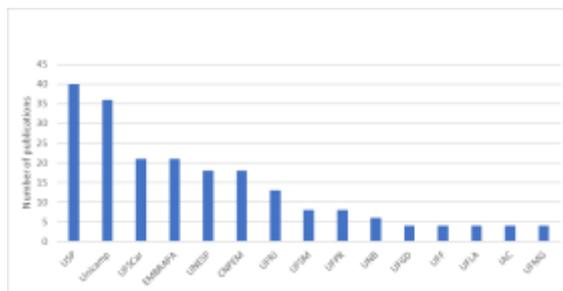
Brazil has a total of 215 publications on the network (5%), of which 42% are in partnership with other countries. The map in Fig. 28 shows Brazil's main publication partners.



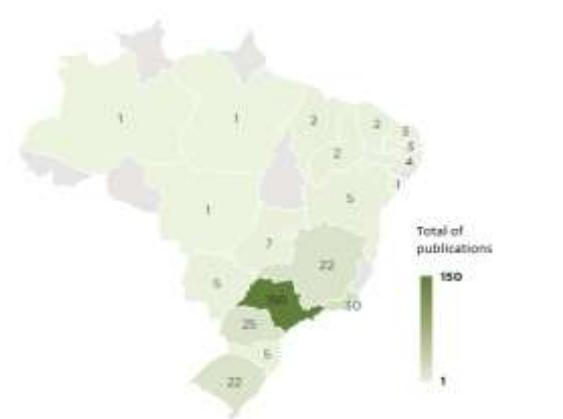
**Figure 28:** Brazil's main publication partners.

Fig. 29 shows the 15 organizations that publish the most in Brazil. It is possible to verify that 12 of the 15 are federal and state universities. It is worth mentioning Embrapa and CNPEM as research institutions that also have a strong role in the subject of bioeconomy.

Fig. 30 shows the map of Brazil with the number of publications by state. These figures make clear the supremacy of the Southeast region, especially the state of São Paulo, in scientific production on bioeconomy. Considering the enormous biodiversity that exists throughout the national territory, this result highlights the need for more STI institutions in bioeconomy beyond the southeastern region.



**Figure 29:** Main institutions that publish about bioeconomy in Brazil



**Figure 30:** Distribution of publications by Brazilian states.

#### 4 CONCLUSIONS

This article presented the results of a semantic similarity network built from a snowball methodology on the term bioeconomy. The results emphasized mainly: the influence of regional characteristics on the types of publications on bioeconomy, the predominance of the biofuels and bioenergy sectors in the subject, and the emphasis on treating bioeconomy through the dimension of biomass or renewable raw materials. The specific results for Brazil showed how the country has been standing out in recent years in terms of the number of publications on the topic. The analysis also showed that research in bioeconomy is still very concentrated in the Southeast region, mainly in the state of São Paulo.

The collected data presents some limitations, such as the use of a single database, the Web of Science database; a methodology based on a small initial sample set (1675 articles); and a partial characterization of the network (15 clusters). Future works will aim to reduce the limitations of the study on the bioeconomy by seeking new data sources, such as other scientific and patent databases, and broadening the scope of the search on the topic.

#### 5 REFERENCES

- [1] CGEE, 2020. Espaço conceitual da Bioeconomia. Brasília: Centro de Gestão e Estudos Estratégicos, 2020.
- [2] European Union, European Commission. A sustainable Bioeconomy for Europe: strengthening the connection between economy, society and the environment. Updated Bioeconomy Strategy. Bélgica: 2018. Available in: <[https://ec.europa.eu/research/bioeconomy/pdf/ec\\_bioeconomy\\_strategy\\_2018.pdf#view=fit&pagemode=none](https://ec.europa.eu/research/bioeconomy/pdf/ec_bioeconomy_strategy_2018.pdf#view=fit&pagemode=none)>. Access in: nov 2020.
- [3] Germany, The Federal Government. National Bioeconomy Strategy. Berlin: jul. de 2020. Available in: <[https://bioeconomie.de/sites/default/files/bmbf\\_national-bioeconomy-strategy\\_en\\_0.pdf](https://bioeconomie.de/sites/default/files/bmbf_national-bioeconomy-strategy_en_0.pdf)>. Access in: nov 2020.
- [4] Nordic co-operation. Development of the Nordic Bioeconomy. NCM reporting: Test centers for green energy solutions - Biorefineries and business needs. 04 fev. 2016. Available in: <<https://www.norden.org/en/publication/development-nordic-bioeconomy>>. Access in: nov 2020.
- [5] United States of America, The White House. National Bioeconomy Blueprint. Washington: abr. 2012. Available in: <[https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/national\\_bioeconomy\\_blueprint\\_april\\_2012.pdf](https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/national_bioeconomy_blueprint_april_2012.pdf)>. Access in: nov 2020.
- [6] Brazil, Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC). Plano de Ação em Ciência, Tecnologia e Inovação em Bioeconomia. 2018. Brasília, DF: Centro de Gestão e Estudos Estratégicos, 2018.
- [7] M. Bugge; T. Hansen; A. Klitkou. What is the bioeconomy? A review of the literature. Sustainability (Switzerland) (2016), v. 8, n. 7.
- [8] D. D'amato; N. Droste; B. Allen; M. Kettunen; K. Lähinen; J. Korhonen, P. Leskinen, B. D. Matthies; A. Toppinen. Green, circular, bio economy: A

comparative analysis of sustainability avenues.  
Journal of Cleaner Production (2017) v. 168, p. 716–  
734.

## 8 ACKNOWLEDGEMENTS

- The Authors are grateful to the technical teams of the CGEE and,
- CGBE/MCTI team of the ODBio project.



MINISTRY OF  
SCIENCE, TECHNOLOGY  
AND INNOVATION

