European Commission

JRC TECHNICAL REPORT

Text Mining for Horizon Scanning

An Insight into Agricultural Research and Innovation in Africa

Borchardt, Steve and Estreguil, Christine

2020

EUR 30349 EN

Joint Research Centre

EU Science Hub
https://ec.europa.eu/jrc

JRC121445

EUR 30349 EN

PDF	ISBN 978-92-76-21446-5	ISSN 1831-9424	doi:10.2760/41501
Print	ISBN 978-92-76-21447-2	ISSN 1018-5593	doi:10.2760/143615

Luxembourg: Publications Office of the European Union, 2020

How to cite this report: Borchardt, S. and Estreguil, C., Text Mining for Horizon Scanning, An Insight into Agricultural Research and Innovation in Africa, EUR 30349 EN, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-21446-5, doi:10.2760/41501, JRC121445.

Contents

Abstract
Text Mining for Horizon Scanning
An Insight into Agricultural Research and Innovation in Africa

This report presents methodological aspects and the results of a horizon scanning exercise where the text mining capacities of the Joint Research Centre's "Tool for Innovation Monitoring" were harnessed to analyse bibliometric data in the field of agricultural research and innovation in Africa in the last five years. The exercise included scanning the institutional landscape and partnerships, resuming key thematic areas, and identifying new emerging fields of research, capturing trends, weak signals and the innovation uptake inside particular topics. The relevance of such exercise is first discussed in the context of enhancing research capacities and strengthening weak institutional networks. Then, a funnel approach is proposed and applied to tackle the broad and multifaceted thematic content of agricultural research in Africa; the value added and limitations of the outcome results are addressed. Ultimately, it is recommended to apply a hybrid approach i.e. a combination of qualitative and quantitative approaches, the latter being based not only on academic records but also enriched with grey literature and real time data media coverage, to implement a horizon scanning exercise.

Aknowledgements
Authors: Christine Estreguil and Steve Borchardt from JRC Unit D.6 Knowledge for Sustainable Development & Food Security

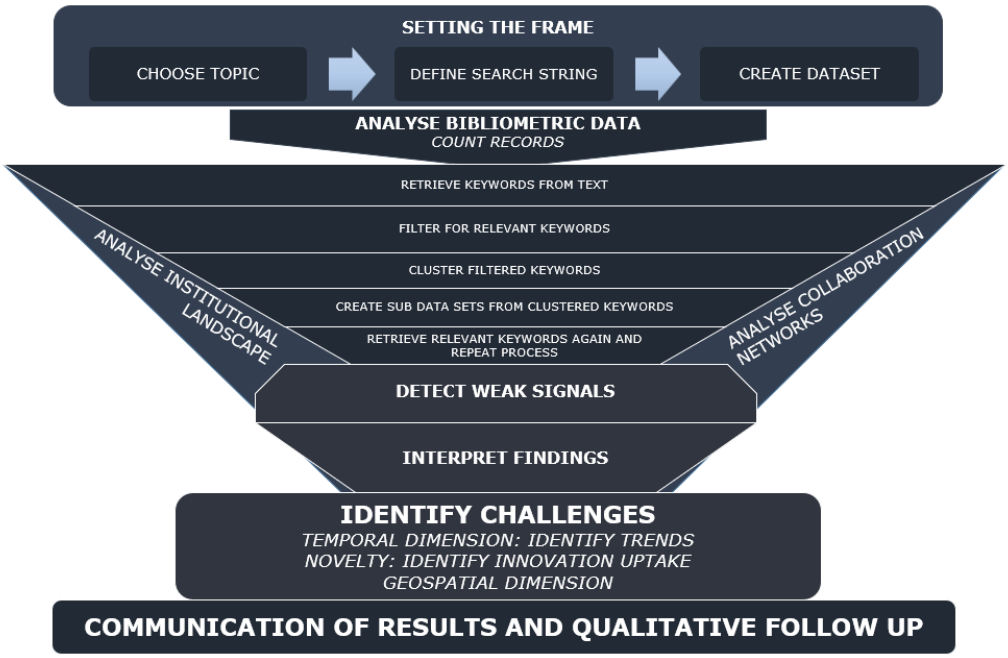
The authors would like to acknowledge and thank Alberto Moro, from JRC Unit I.2 Foresight, Modelling, Behavioural Insights and Design for Policy, for his support and availability during this study, and François Kayitakire, from Unit D.5 Food Security, for his valuable comments to improve the report.

Executive Summary

Constant changes, upcoming trends and demanding challenges create dynamic interactions in our highly complex systems on all levels – be it through the development of impactful innovations, the adoption of disruptive technologies or newly emerging environmental problems or societal crises. The strong interconnectedness of our systems further amplifies those dynamic interactions. In this context, Horizon scanning and foresight tools and methods increasingly gained popularity and relevance across various sectors and among different stakeholders as they aim to support decision makers in better anticipating those trends and changes in a timely manner so that adequate responses and solutions can be developed.

Originally, foresight and horizon scanning activities were largely carried out by participative approaches with experts from diverse fields and with different expertise. Discussions and workshops were occasionally stimulated by the outcomes of literature reviews. However, new sources of information are rapidly emerging, featuring a vast amount of different types of data that become nearly impossible to get hold of. As this data can provide eminently valuable, timely, exhaustive, and impartial insights for horizon scanning and foresight activities, new automated desk-based methods utilising quantitative data analysis approaches are gaining more and more relevance in the field.

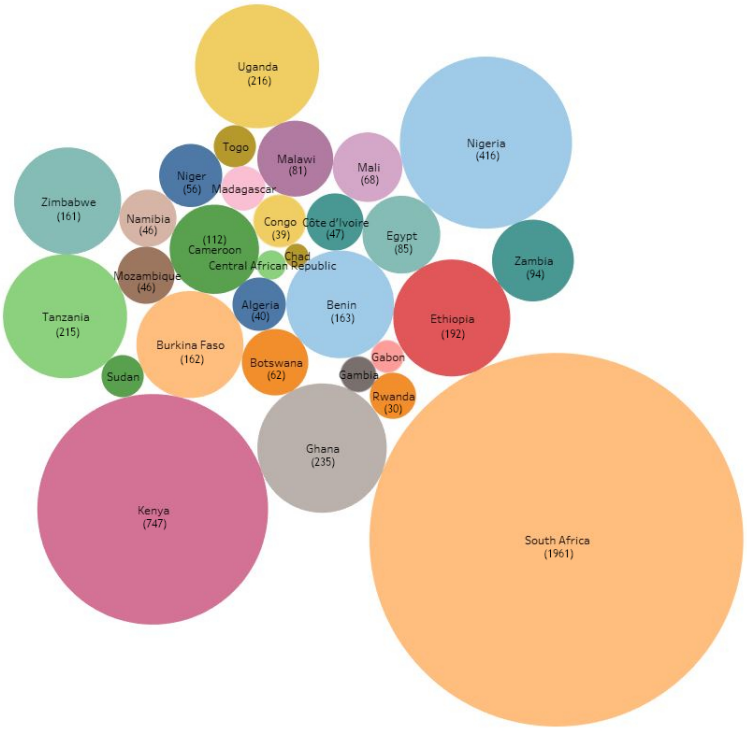
This report presents methodological aspects and the results of a horizon scanning exercise where the text mining capacities of the JRC's "Tool for Innovation Monitoring" were harnessed to analyse bibliometric data in the field of agricultural research and innovation in Africa in the last five years. The exercise included scanning the institutional landscape and partnerships, resuming key thematic areas, and identifying new emerging fields of research, capturing trends, weak signals and the innovation uptake inside particular topics. TIM focuses on academic records and patents. It allows to create customised data sets through individually defined search queries that derive text meta data from different sources: Scopus, the European Patent Office and EU projects. These datasets can then be analysed according to their most relevant keywords in order to better understand and structure the current knowledge base and also detect weak signals within the horizon scanning process. For the thematic clustering, a fit-for-purpose process was designed as a funnel approach. As the first keywords retrieved from the dataset do not necessarily provide the level of detail needed for identifying weak signals, these keywords are filtered, clustered, and used as input for defining new data sets. The process of keyword retrieval is then repeated for each of these sub data sets to dive deeper into the bibliometric data to eventually detect relevant signals. In a last step, data sets for each signal are created to further contextualise the signals and gain a better understanding of aspects associated to the weak signal. Regarding the institutional landscapes and stakeholder constellations, TIM creates illustrative network graphs showing connections between countries or organisations based on author affiliations of a publication record in the data set (see figure below for key steps of the analytical process).



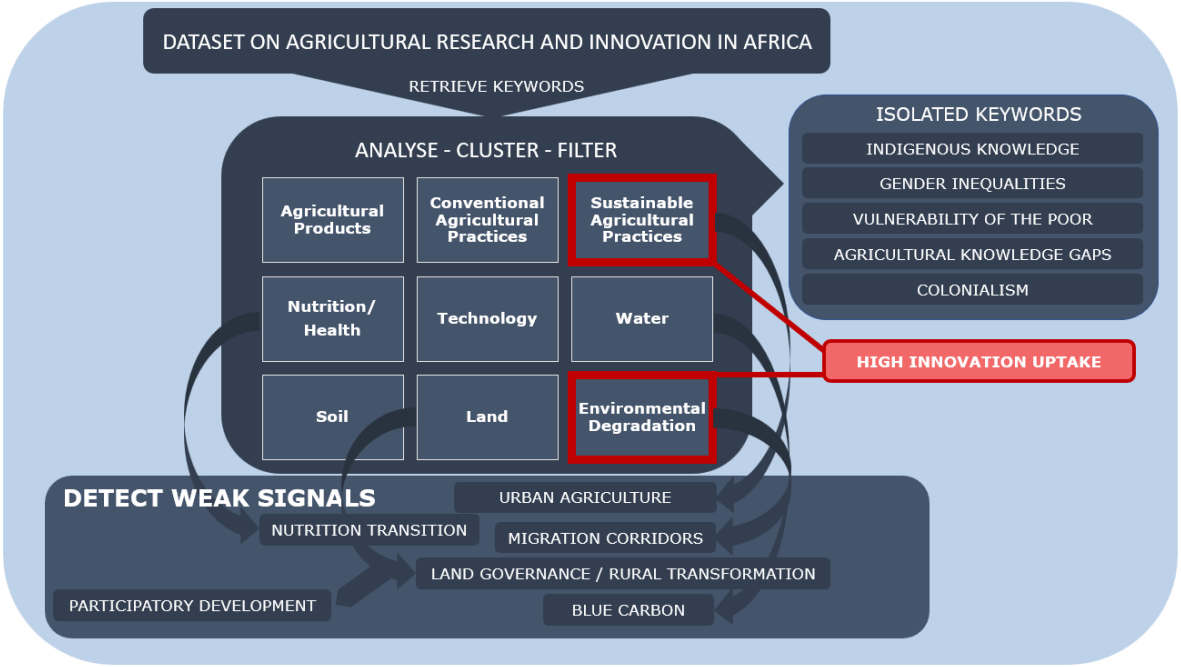
Key steps of the analytical process

The method was demonstrated for the field of agricultural research and innovation in Africa. First, scanning the African research and innovation provides a valuable synoptic insight on the current institutional landscapes and networks, particularly in terms of most participative countries such as South Africa, Kenya and Nigeria (see figure below), best performing institutions (mainly South-African ones), main active partnerships and collaboration axes (mainly with Europe and United States). Policy makers, such as the European Commission' s Directorate-General for Research and Innovation, could use this information to design research programmes that use in an efficient manner the existing networks and identify weak connections in need to be strengthened. Improved understanding of the structure of these networks can help to boost research capacities within Africa and not only within consolidated networks that already exist. The exercise is relevant to target those countries acknowledging weak participation for new research calls, emulate best practices in those countries and monitor their uptake in the academic field.

Further, the analysis led to the identification of nine thematic clusters that featured various weak signals. The clusters addressed different aspects ranging from conventional and sustainable agricultural practices, specific agricultural products and threats to them, to questions of nutrition, land, tenure rights and rural electrification as well as issues related to technology (particularly remote sensing, telemedicine and e-governance), available natural resources (particularly water and soil related), their management and their degradation. Among these clusters, different weak signals were detected like nutrition transition, green manure, blue carbon, participatory development, land governance, urban agriculture, and rural transformation. Additionally, few isolated still important aspects could not be assigned to clusters, like issues of gender inequality, the persistence of colonialist structures or the lack of social capital and the need for capacity building. Finally, clusters like "Sustainable Agricultural Practices" and "Environmental Degradation" followed by water and technology feature the highest share of innovation in the data set.



Institutional landscape of African research and innovation: importance based on number of publication records per country



Thematic content analysis:
case study on agricultural research and innovation in Africa

To conclude, the funnel approach proposed in this report seem to be most suitable for tackling broad and multi-faceted issues and for contextualizing obtained keywords in a comprehensive way to gain more insights on the bigger picture. Nonetheless, some open questions remain like why allegedly expected keywords were missing, how to solve the predominance of hot topics of the recent past as dictated by the publication timeframe

(2014-2019) and how new urging matters can be detected rather than only consolidating long-standing well-known issues. The presentation of the thematic clusters also showed the difficulties that potentially arise when analysing them as all of them are multi-dimensional in their nature and feature a high degree of complexity. It was concluded that experts with specific expertise in each of the clusters could greatly contribute to the qualitative interpretation and contextualisation of these quantitative results. Overall, guidance on development projects, management decisions and policies can be provided if experts utilise the tool and its outcomes in the right way.

The source and the time dimension of the data used as input in TIM is even more important when there is the need to capture 'unknown' topics and technological and innovative issues in which the agricultural sector in Africa could shift to; two ways forward were identified, (1) start the analysis with a focal point on known topics that have been rarely researched to highlight sub-topics with a high potential and identify institutions engaging in research activities related to those topics to show where potential innovators might be located, (2) include recently available grey literature where articles and research reports that are not referenced in journals are available from many researchers interested in African agriculture. Ultimately, recalling the purpose of horizon scanning as a first step to engage on strategic foresight and supporting anticipatory governance, these results need to be communicated appropriately and disseminated in a way that policymakers and other relevant stakeholders consider them respectively within their dialogue and decision-making processes. While new data sources are constantly emerging along with powerful tools and methods to analyse them, this exercise showed not only the great potential for horizon scanning laying within this field, but also the necessity to involve experts in the process to first precisely define the scope of the search particularly for multi-faceted complex topics, and second, evaluate the outcomes accordingly and make the results of a horizon scanning count.

A hybrid approach, i.e. a combination of qualitative and quantitative approaches, the latter being based not only on academic records but also enriched with grey literature and real time media coverage, seem to be the most efficient and persuasive way to implement an horizon scanning exercise.

1 Introduction

The increasing degree of complexity in our globally intertwined systems and the resulting interactions of different components inside these systems lead to dynamic situations and changes that can be accompanied by new trends, challenges and technological innovations with far reaching implications for society, economy and our environment. horizon scanning can serve as a tool that enables decision makers to be aware of these trends in a timely manner to adequately respond to them and consider them respectively in their decision-making processes. Therefore, horizon scanning can be seen as a search or scanning process at the verge of our known environment, potentially even reaching beyond its borders in order to detect emerging events, trends or developments which may present opportunities or threats to our society, economy or environment.
The ever-increasing amount and diversity of easily accessible data provide an invaluable source of information and great untapped potential for activities like a horizon scanning. As data analysis, data processing and analytical methods, tools and techniques constantly evolve, new possibilities open up to design a horizon scanning process by utilizing quantitative data and methods. Though these tools and methods can help to get hold of the escalating amount of data and contribute to further automating specific analytical steps, outcomes still need to be evaluated and processed by experts who are able to fine tune and to adjust the right screws in the underlying procedures. A close interaction between horizon scanning experts and the policymakers needs to be ensured during the entire process to foster stakeholder engagement that can contribute to translating resulting recommendations into tangible actions.

Differences between horizon scanning and foresight activities
Horizon Scanning can be considered a methodological tool or instrument to gather knowledge intelligence, get insight on a topic of concern and prepare for actions in the future (Figure 1). It is often embedded in the first phase of a foresight process (Figure 1, 2). Foresight has become an indispensable tool in policy-making processes to anticipate governance. It aims at first understanding “the picture of now” and capturing ‘pockets of future’ in the present time, to then propose to policy makers actionable knowledge and pathways to influence the present, help them defining priorities, setting strategies and building capacity. Foresight involves participatory sense-making sessions with experts acting as a think tank to identify and analyse signals, translate horizon scans into actions, help defining their time horizons, and finally set objectives and targets. In its last phase, foresight includes a social process to facilitate dialogue and develop social learning across experts and multiple stakeholders from different background and sectors. Learning between all participants in the process is an important outcome, and for example can be built up through the exchange of scientific facts and data evidence, expertise, and traditional knowledge.

Figure 1: Nine step process of a horizon scanning Exercise. Garnett, K. 2018.

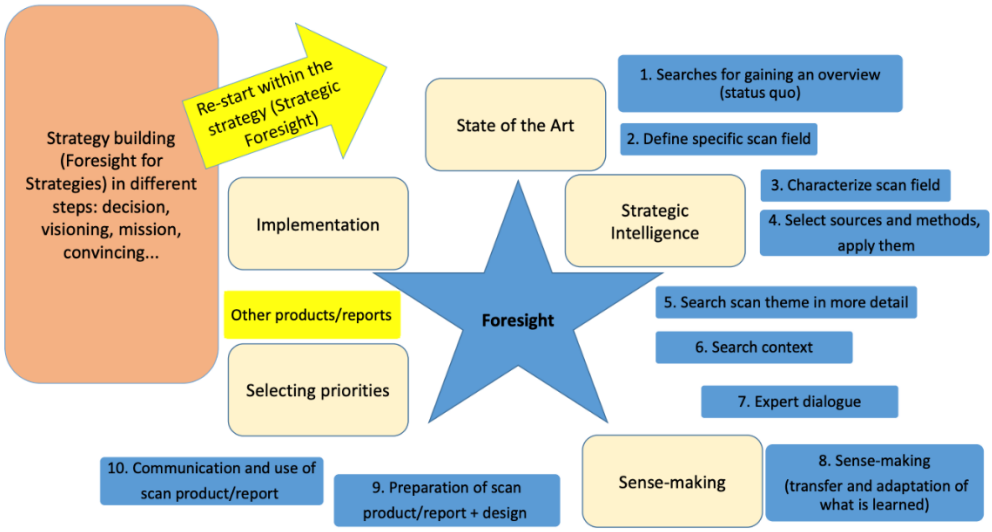


Figure 2 Steps of horizon scanning (in blue shade) integrated within a foresight process Cuhls, 2019.

There is often confusion and uncertainty on how foresight and horizon scanning activities are interlinked with each other and what their actual relation is. While both target the same outcome of better anticipating emerging trends, technologies and challenges, practical and theoretical differences exist. The majority of researchers and practitioners describe foresight as a comprehensive toolbox comprised of different activities that strive to include various stakeholders and that are more process-oriented than a horizon scanning. Horizon scanning can be seen as a search process where signals are detected and identified. However, this search process does not necessarily include sense-making or interpretation of results and neither requires the translation of signals into tangible actions for implementation. Furthermore, it does not have to be designed as a participatory process. On the contrary, foresight is often applied for developing strategies within decision making processes or to advice policymakers by utilizing discussion and dialogue formats.

Horizon scanning is not directly part of a strategy building process, but it can be incorporated in the beginning of a foresight process (Figure 2, Cuhls 2019). These possible connections combined with a lack of conceptual clarity have led to theoretical confusions that also percolated into practice where foresight activities were being applied. Within the sphere of horizon scanning, activities have broadened with numerous applications in different fields that increasingly make use of automated approaches, providing results that are valuable and vague at the same time (Cuhls et al. 2015). Thus, horizon scanning should not be deployed as a standalone exercise, but rather embedded within strategy building activities that translate the generated insights into actionable recommendations providing the knowledge base for decision making processes.

This report focuses on horizon scanning in the first phase of foresight activities and strives to illustrate the potential benefits of utilizing text mining tools for a horizon scanning exercise within a desk-based approach. It aims to showcase potential stumbling blocks when designing and conducting a horizon scanning exercise in general and with special regard to quantitative data analysis. The methodological approach is explicitly outlined, thus providing a potential blueprint that can be applied in other contexts as well. The horizon scanning exercise serving as a base for demonstration is conducted in the context of Research and Development for the sustainable use of natural resources in Africa; it addresses issues of active participation and partnership as well as thematic clustering including innovation uptake. An insight is provided for the agricultural domain. Finally, the results are presented, and lessons learned discussed along with an outlook for the way forward in the future of designing horizon scanning and foresight activities.

2 Text Mining for horizon scanning

Continuous streams of data are being generated in various formats, presented through different mediums, and coming from multiple platforms in consistent time intervals, providing an overwhelming source of information on what is happening in our world. These sources can be useful when conducting a horizon scanning exercise, but manually processing the sheer amount of incoming data in time becomes impossible. This is why efficient data analysis techniques emerged that can be deployed in order to get hold of the constant streams of information in a more automated way. With regard to horizon scanning, text data are highly relevant thanks to the great density of information that is presented in a very timely manner and text mining techniques are valuable to assist in processing a lot of data in a short period of time. Processing the text and filtering it for relevant information, performing cluster analysis based on this information as well as content analysis and comprehension can be useful text mining activities for horizon scanning.

2.1 Introducing the JRC's Tool for Innovation Monitoring (TIM)

The '*Tool for Innovation Monitoring'* (TIM)[1] developed by the Joint Research Centre (JRC), the European Commission's science and knowledge service is combining a set of text mining steps into one tool, allowing for an analysis of text data from various sources that feature relevant information in the fields of research, innovation, technology and development. TIM was specifically designed to provide European policymakers with a better understanding about the dynamics of innovations so that European policies can be designed appropriately to foster innovation. Trends and emerging technologies in research and industry can be tracked with TIM, relevant innovation actors or regions of particular importance for certain technological or scientific areas can be detected and highlighted along with the emergence of the research networks in which they are embedded. Furthermore, gaps within research and technology can be identified in order to allude to potential investment needs.
Within TIM, users can create their own data sets by defining search queries that allow for great flexibility during the design process in order to target topics specifically, so user needs and interests are met accordingly. These queries can be limited to particular time periods, to particular countries or institutions and can be combined through logical operators in order to produce accurate search results (see chapter 2.4 for more information).
Once a data set is created, the data can be analysed in various ways. TIM provides an overview about the data set specifications in which the number of publication records sorted by different data sources and data types as well as the emergence of publication records over time can be analysed. Based on the author affiliations of each publication, networks are created and visualised in TIM, that allow users to analyse collaboration axes, partnerships between countries and the institutional landscape related to the data set in general (see chapter 3.2 on institutional landscapes and partnerships for more information).
TIM automatically retrieves keywords that are provided within the meta data of each publication record. Furthermore, the tool applies different text mining and processing algorithms on the available text data to automatically identify relevant keywords. The keywords provided in TIM build the basic input for the horizon scanning, as potentially interesting signals can be found within them. We anticipate that the screening and

[1] EU, Joint Research Centre, TIM Analytics, http://www.timanalytics.eu/

evaluation of keywords as well as the making sense of all retrieved results will not be fully automated and will require the supervision of experts.
The tool is able to create networks graphs out of keywords in each publication, for example in figure 3, showing an example network of a sub-dataset in the context of African agriculture that illustrates most relevant keywords derived from the data. The strength of the relationship between keywords in the network is determined by the frequency in which they are mentioned together in a publication, whilst the size of each node relates to the overall frequency of the keyword in the literature dataset. All of the above-mentioned data can also be exported from TIM as Excel-tables or as Gephi network files (GEXF).

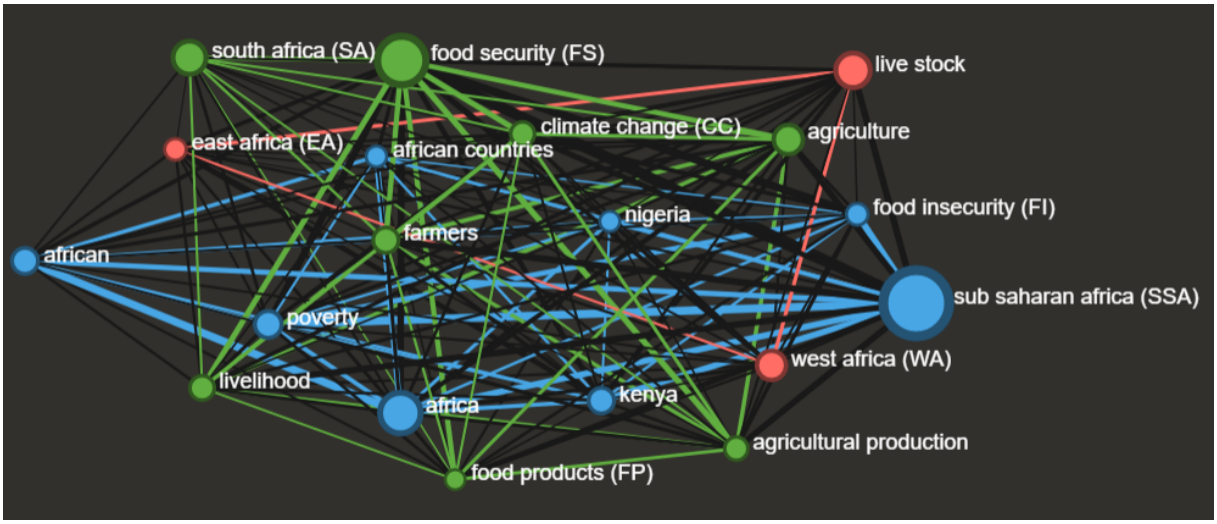


Figure 3: Example Network of Keywords visualized in TIM (colours of links based on automatic clustering algorithm in TIM – can be neglected).

All in all, TIM taps data sources that can be highly relevant for horizon scanning and allows for creating data sets to which insightful bibliometric text analysis techniques can be applied in order to generate an output that can be utilized for the detection of upcoming trends, challenges or other signals relevant to a particular topic – all bundled in a convenient user interface.

2.2 Data Sources

Not all sources of text data can be equally utilized for a horizon scanning exercise and the selection of data sources also largely depends on the scope of the exercise itself. When focusing on technological aspects and innovation, texts from the news and media may not provide the information base needed, whereas societal changes and citizens' concerns might be well reflected in these sources. Reversely, the same applies for academic literature.
TIM considers data from three different data sets: academic literature from Scopus, European patent data from the European Patent Office's PATSTAT data base and information from the Cordis data base, containing information on projects that were funded by the EU's framework programmes (FP5, FP6, FP7, Horizon 2020) for research and innovation since 1998. Since Scopus is the biggest data base among the three, data sets that are created in TIM are usually dominated by Scopus data, while European patent data and Cordis provide only a small share in the data sets, though this also depends on the scope of

the data set and a more specified query can potentially lead to more patent data or Cordis records featured in the data set (see figure 4 below).

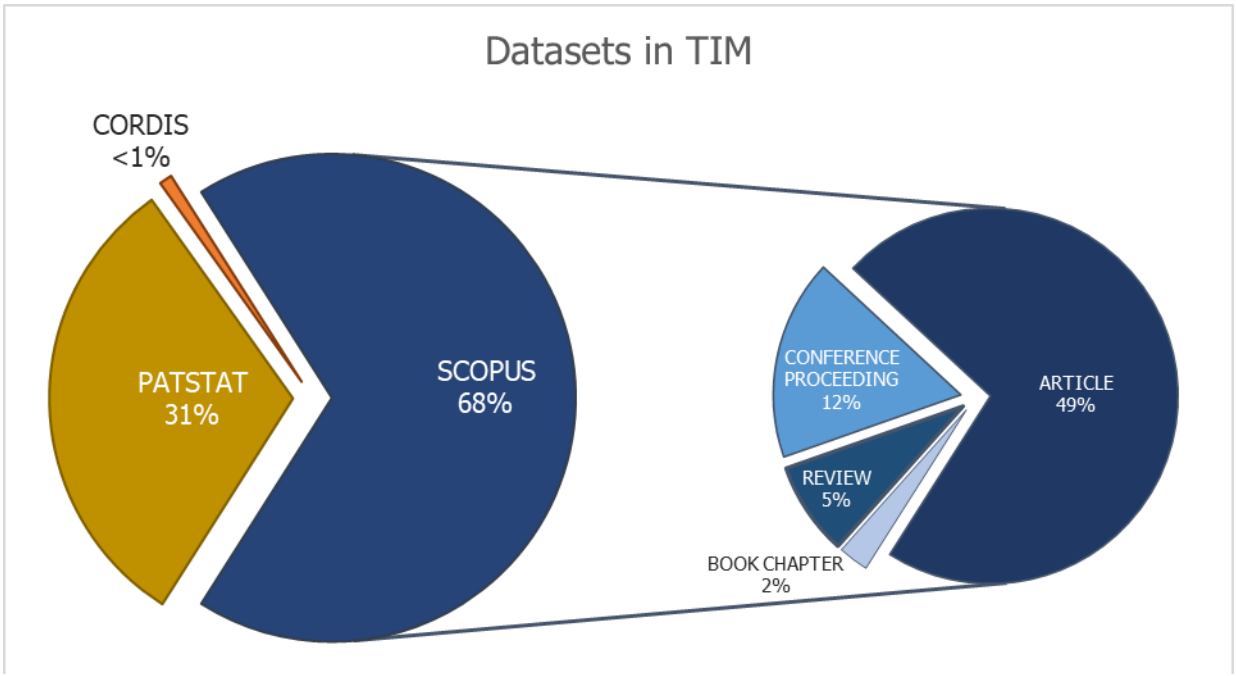


Figure 4: Number of records in TIM grouped by different data sources.

Scopus

The Application Programming Interface (API) that is utilized in TIM to access the Scopus database allows to retrieve different meta data of a publication along with its abstract. The database includes articles that can either be original research, opinions, or Article-in-Press. Additionally, Scopus is complemented by conference proceedings (conference abstracts, conference summaries or even full papers), reviews (including short surveys) as well as book chapters.

PATSTAT

The PATSTAT database covers patent data coming from more than 90 patent authorities with documents dating back until 1996. Patent documents are grouped as a patent family if at least one of the documents is written in English. A patent family is assumed to account for one invention. Since patent documents are only published 18 months after their application, the data base itself is updated two times per year. Cleaning and processing of the patent data coming from sources other than the European Patent Office takes more time and there is a significant lag in the data of the PATSTAT database for the last three years. This lag should be kept in mind, especially when conducting a horizon scanning exercise for which the more recent years are usually the most interesting ones as well.

CORDIS

The CORDIS data set, accessible through the European Union Open Data Portal, features a subset on granted EU projects that receive or received funding under the various European Union's framework programmes for research and development, starting with the Research Framework Programme FP5 from 1998 to 2002, followed by FP6 (years 2002-2006), FP7

(years 2007-2013), Horizon 2020 (years 2014-2020) and eventually by Horizon Europe (2021-2027) in the near future. For each project, the data set provides information on the project objectives, subjects, participant countries, reference, starting date, title and programmes that can all be used for text mining analysis.

Relevance and timeliness of data sources

The three above-mentioned data bases are complementary in providing an overview of the academic field by encompassing a broad range of research topics, different methodological approaches, and applications across geographic and thematic lenses, and by scanning partnerships and identifying innovations. They undoubtedly provide a valuable source of information for a horizon scanning exercise to capture emerging trends, technologies, innovations, or challenges.

The timeliness of the data - essential for horizon scanning - remains an issue of concern. As mentioned previously, there can be a significant delay between knowledge generation and the date of publication and dissemination (as indicated by PATSTAT data lag, time consuming process of publishing an article in an academic journal or the time when an application for EU funding of research is written, submitted, processed and approved). Alternative more qualitative approaches than TIM, such as expert group workshops, also rely on similar sources of knowledge, so the concern of timeliness partially remains. The inclusion of the so called 'grey literature' offering more recent data that are not scientifically peer-reviewed, either generated in projects (e.g. for industry) or in the news could be one way to improve timeliness. In such case, a compromise between reliability (also accountability) and timeliness of the data needs to be found (see chapter 5 for more information).

2.3 Methods

Text data is usually provided as unstructured information, making it thus necessary to apply text mining techniques in order to extract meaningful information from the text. While unstructured text can usually be processed without great efforts by humans, it can become a much more challenging tasks for machines. Available fit-for-purpose methods and algorithms cover various topics ranging from data mining, natural language processing and information retrieval to machine learning and deep learning approaches (Allahyari et al. 2017b). The report focuses on the methods utilised for the present horizon scanning exercise – information retrieval and text summarization.

Information Retrieval

Information retrieval is the initial step when conducting a horizon scanning exercise through quantitative text data analysis. Based on an articulated information need, relevant information within unstructured data sets is identified and accessed (Faloutsos and Oard 1995). A widely adopted field of application where information retrieval is implemented can be found in search engines that are designed to provide users with information coming from an unstructured massive data set – the internet. With the application of web search engines, the field of information retrieval managed to shift from solely being used in academia where it was applied to navigate through library records and academic publications to becoming the preferential underlying mean for most people for finding and accessing information (Manning et al. 2008).

For the present exercise, this information need was articulated in TIM as a conditional search query for documents within the three previously described data sources. Additionally, TIM utilizes information retrieval to access the meta data of each record in the data set. In this way, information on keywords provided by authors can be accessed to get a better idea of what a record in the data set is referring to. In a next step, retrieved meta data can be used for further quantitative analysis.

Text Summarization

Information retrieval is an essential entry point, but it does not include analysing and making sense of the retrieved text data. These tasks are deployed thanks to text summarization techniques. Text summarization can either be performed by extracting information units from the original text (extractive summarization) or by synthesizing information in a way that may not be found in the original text (abstractive summarization) (Allahyari et al. 2017a). Abstractive Summarization is based on complex language modelling, usually utilizing machine learning and neural network approaches to create these target texts by generating new phrases or rephrasing words. This is a challenging task that is highly complex, which is why the majority of studies and work on text summarization focuses on extractive summarization (Lin et al. 2018). Accordingly, TIM does not include any features on abstractive summarization and is focusing more on the extractive summarization approach by identifying relevant keywords that summarize a document and the entire corpus (collection of documents). However, results from an abstractive text summarization process could potentially also offer interesting insights on the content of the data set (see section 5 on conclusions).

The relevance is quantified by using the TF – IDF algorithm (Term Frequency – Inversed Document Frequency). The first part on term frequency involves checking how many times a term is present in a given document. Several pre-processing steps are applied to the text, such as tokenization, the removal of special characters, lowercasing all words and stemming. When tokenizing (i.e. segmenting) a text or a document, the text is divided into different characters, either split into separate words or separate sentences. After tokenizing the words, the other steps can easily be applied, and the final step of stemming transforms the words to their stems in order to standardise the terms appropriately for the frequencies that are being calculated subsequently. As the length of documents in a corpus usually varies, the term frequency is rectified by dividing the frequency of a term by the total number of terms in a document. The Inversed Document Frequency is the second part of the algorithm and describes a method of putting a weight to each term in order to account for the importance of terms in the overall corpus and to “mute” stop words like “and”, “to”, “of”. Consequently, generic words that occur more frequently, receive a lower weight and less frequent words are assigned a greater weight. In a final step Term Frequency and Inversed Document Frequency are multiplied to calculate the final TF – IDF scores where terms with higher scores can be considered to be more important (Qaiser and Ali 2018).

In TIM the described steps are applied on the data set in order to obtain the most relevant keywords. One additional step is carried out when calculating the TF -IDF scores in order to consider n-grams as well, because otherwise, relevant keywords consisting of more than one word (e.g. climate change) are being neglected.

Network Analysis

Although not a text mining technique in itself, network visualizations and analyses are performed in TIM on the basis of the output of text mining. Network graphs can provide valuable insights for horizon scanning on the institutional landscapes and partnerships around a particular topic. In TIM, complex connections and relationships can be visualized and clustered by performing network analysis that can be carried out for topics, countries, regions, or institutions (Joanny et al. 2015). Furthermore, temporal filters can be applied to the data set, to see how partnerships and organizational networks develop over time, who are the key stakeholders and if newly emerging organizations or institutions enter into play. The networks are created by retrieving the author affiliations of each publication record and creating links between organizations (see section 3.2.1) or countries when their authors collaborated on a publication, a patent or within an EU funded project (see figure 5 below for an example).

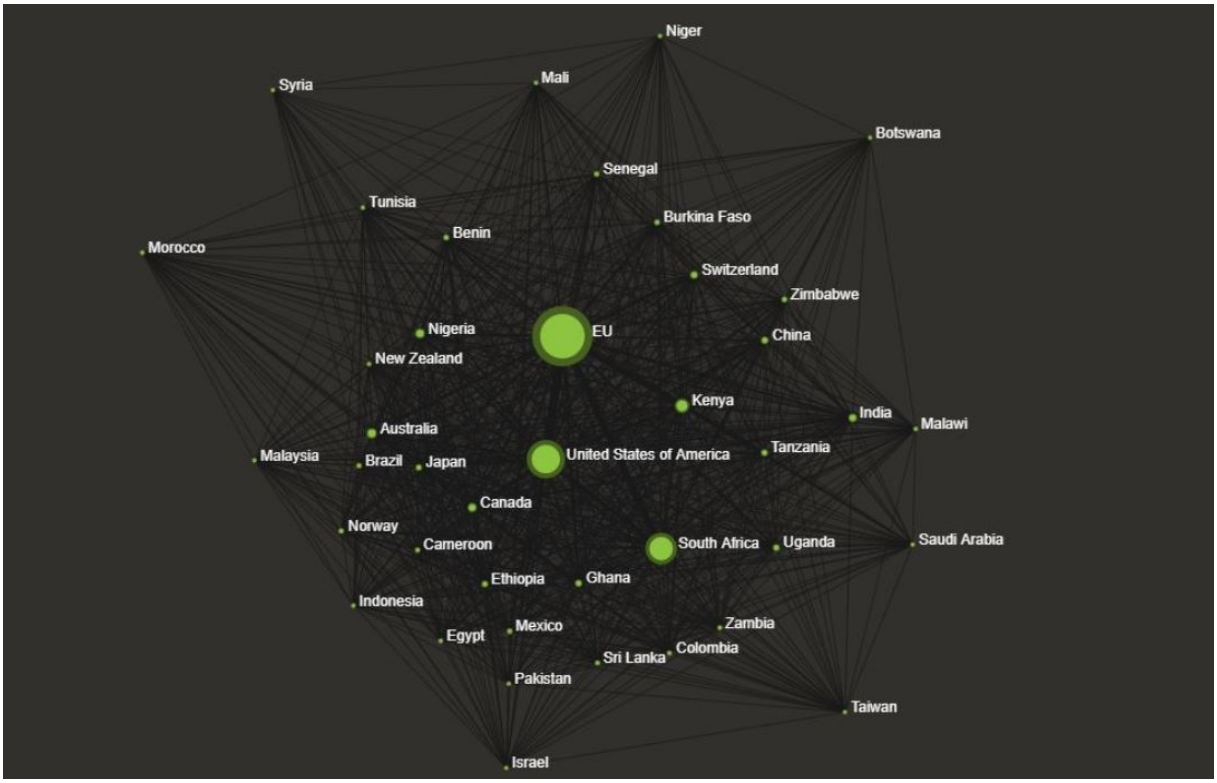


Figure 5: Example of Network Visualisation in TIM.

Two shortcomings of the current technique are worth mentioning from previous exercises; *(1) The nodes correspond to the frequency of the terms in the dataset – meaning the bigger nodes represent more frequently occurring terms in the dataset. The node sizes can only be compared within a given dataset, as the size is assigned in relation to frequencies of the other terms inside the same dataset and not in absolute matters and thus, node sizes across different datasets cannot be compared to each other (Boelmann & Telsnig 2016).*
(2) all networks in TIM can also be exported to Excel as tables showing nodes and edges and even providing network analysis measures like closeness centrality, harmonic closeness centrality and betweenness centrality. These exported tables listing the keywords ordered by their frequencies within the dataset serve as main input for the scanning process to detect weak signals (keywords with higher term frequencies respond to bigger node sizes in the network and thus, are listed at the top of the exported Excel tables). TIM takes keywords

provided by the authors of a particular publication and cleans them by accounting for capitalisation, hyphenation and spelling. However, the tool reaches its limits when it comes to discerning between conceptually overlapping or similar terms that can potentially be used interchangeably but are semantically different from each other (for instance aquaculture and fisheries management are highly related with each other, but will be detected separately due to their semantic differences). Thus, expert judgments for evaluating the retrieved keywords remains a necessity (Boelmann & Telsnig 2016).

2.4 Designing a horizon scanning exercise in TIM

This section outlines how all of the above-mentioned steps and techniques come together to design a horizon scanning exercise that utilizes TIM capacities.

Critical Steps of horizon scanning

Before describing the horizon scanning process, some conceptual aspects need to be clarified. Horizon scanning has two key functions: an alerting function that shall enable policy and decision makers to better anticipate upcoming issues and a creative function that facilitates the development and implementation of solutions and new ideas based on the processing, analysis and integration of horizon scanning results. These results usually manifest as weak or early warning signals that need to be detected. Detecting those signals can be difficult as they are characterized by incompleteness and uncertainty. In a next step, emerging issues can be identified when different weak signals are detected and can be combined on the basis of their possible interconnected impacts. This step requires an accurate assessment of the spotted weak signals, including gathering of signals, clustering them and assessing their significance and framing the topics around clustered signals. Once significant emerging issues have been identified, they need to be analysed and interpreted with regard to their policy relevance in order to eventually translate those findings into appropriate and convincing policy recommendations. The timing for disseminating these recommendations and the information on detected signals as well as the communication layout are crucial for their consideration in policy and other decision-making processes, as outcomes usually have to fight for the attention of stakeholders if they do not align with their interests and ideas.

For the scanning process, two different approaches exist – issue-centred scanning and exploratory scanning. Issue-centred or sectorial scanning strives to gather core documents on a certain topic or sector (e.g. particular policy domain) and to scan those documents for weak signals in order to create a storyline that is based on the future narratives outlined in the respective documents. Exploratory scanning seeks to assemble possible emerging issues from various data sources without providing analysts with a predefined framework or focus. The approaches are not mutually exclusive, and can be combined, for instance by starting with an exploratory scanning to generate a hypothesis that can then be evaluated with the help of an issue-centred scanning approach. This combination of approaches is illustrated in figure 2 for a strategic foresight process. These scanning processes can be greatly supported by new data analysis methods like text mining (Amanatidou et al. 2012).

Process Design

The present exercise started with implementing an exploratory scanning approach on a very broad topic (sustainable use of natural resources). Setting a frame provides a minimum degree of guidance for the process and proved necessary to be able to define a search query. Within TIM, the user has to first create a data set that is based on a search query – targeting on specific topics that can be combined through logical operators, restricted to particular time periods and even to affiliations e.g. publications made by selected organisations only.

The construction of the search string was the most crucial and time demanding step for obtaining a data set which meets the user needs. The search string was composed of four different blocks: one search block on the topic itself, one on horizon scanning (capturing words describing innovation, novelty, upcoming trends etc.), one with the geographical scope (Africa) and one on the time period (restricting the entire data set to the years between 2014 and 2019).

The blocks were then combined through logical operators, as were the terms inside the search blocks (see footnote[2] below for final search string related to the agricultural domain). Though it seems to be an allegedly trivial step in the process, it is one of the most impactful and simultaneously difficult parts, as the choice of keywords defines which publication records go into the data set and thus, determines which content is being scanned. Keywords need to be carefully selected in order to ensure that all different aspects are captured appropriately. The only search block where certainty of intended results is ensured is the block defining the time period. However, for the other search blocks it is much more complicated to find all the keywords referring to either the specific topic, the geographic location or the horizon scanning itself and most certainly, a bias produced by the individual perspectives, perceptions and knowledge of the "Horizon Scanner" is being introduced at this point. Even though the validity of the dataset is subsequently checked, it can only be verified whether the dataset captures intended records, but not whether all relevant records are being included in the data.

A complementary way of providing relevant keywords consisted in calculating term frequencies in the documents in relation to the frequencies of the overall corpus.

A data set was created in TIM according to the conditional search string and cleaned if necessary. For this step, it is important to reach a statistically significant number of documents (at least a couple of thousand records) to ensure that the applied text mining techniques produce reliable outputs that are statistically sound. After creating the data set, the included documents should be screened through their titles to verify that the search string led to the desired documents. The compliance of the data obtained need to be checked against the conditional search rules, which are then adjusted in case of non-compliance.

---

[2] *Search String in TIM: ti_abs_key:(("food security" OR "food insecurity" OR "cropland" OR "agricultural practices" OR "agriculture" OR "small scale food producer" OR "small scale farmer" OR "sustainable agriculture" OR "soil fertility" OR "soil permeability" OR "soil sealing" OR "soil quality" OR "access to land" OR "land ownership" OR "land conflict" OR "tenure right" OR "family farming" OR "agricultural productivity" OR "organic agriculture" OR "organic farming" OR "food production" OR "food accessibility and equity" OR "food and feed" OR livestock OR nutrition OR hunger OR "urban farming" OR "crop pest" OR "crop disease" OR "yield production" OR (yield AND "weather patterns ") OR ("fair benefits" AND "food chain") OR "climate smart farming" OR "sustainable intensification" OR "alternative agricultural production methods") AND (future OR emerging OR innovative OR disruptive OR visionary OR exploratory OR unexpected OR novel OR novelty OR innovation OR disruption OR "cutting edge" OR "latest development" OR "new technologies" OR "earth observation" OR "remote sensing" OR GIS OR geospatial OR geointelligence OR "cross-cutting" OR vision OR trends OR trajectory OR change OR projections OR foresight OR "systemic innovation" OR "innovation ecosystems" OR "cross-sectorial innovation" OR "eco-innovation" OR "societal challenges" OR "trans-disciplinary" OR "digital revolution" OR "open data") AND ("africa" OR "african countries" OR "north african" OR "east african" OR "west african" OR "sub saharan" OR "central african" OR "African Union" OR REC OR RECs OR "Regional Economic Communities")) AND emm_year:[2014 TO 2019]*

This step also alludes to one of the most important aspects of this exercise – the fact that the construction of the search string determines which documents are retrieved and thus, ultimately which results are being obtained (see chapters 5 and 6 for more information). As trivial as it may sound, this is something that always needs be bore in mind, especially when conducting a horizon scanning exercise that should ideally strive to search for the "unknown unknowns". This bias that is being introduced immediately at the beginning of the exercise has to be considered when interpreting intermediate or final results at any step of the process (see chapter 5 for suggestions on how to solve this issue).
All of the above-mentioned data can be exported from TIM as Excel-tables or as Gephi network file (GEXF). In particular the keyword tables serve as a valuable entry point where trends and signals can be detected and analysed. After extending user capacities in TIM, the most common keywords were retrieved – one table with the top 300 keywords provided by the author(s) and one with the top 300 keywords derived from the previously described TF-IDF calculations. These tables were then screened and filtered for potentially useful signals. The screening and evaluation of keywords cannot be fully automated and require the supervision of experts. After the qualitative expert evaluation, the retained keywords were clustered by topic to better structure the output and get a better overview on sub-topics.

In a next step these keywords were then used to create new search strings with which new sub data sets in TIM were created. TIM retrieves also the temporal profile of the data, allowing for assessing the emergence of a topic for a specific period of time relative to the other sub data sets or their relative importance in general when comparing the number of publication records across the data sets (see results chapter 3). Afterwards, the routine from the first step is repeated – author keywords and TF -IDF derived keywords are retrieved, exported, screened, and evaluated again. At this point, keywords may already contain potentially interesting signals that are more in line with what a horizon scanning strives for – the detection of weak signals.
These allegedly weak signals are then used again to create new sub data sets in TIM and the whole routine is being repeated. The approach can be illustrated like a funnel that is applied on a broader topic and through retrieving and evaluating keywords, the bibliometric data is being filtered and narrowed down until weak signals can be detected. The above-mentioned significant number of documents cannot be ensured anymore after the sub data sets are being created since their scope is already quite specific and thus, the number of publications can be relatively low. However, a statistically significant number of records is only necessary in the initial stage of the process because the derived keywords from this stage will then provide the input for the subsequent steps. The figure below summarises the approach of the horizon scanning exercise.

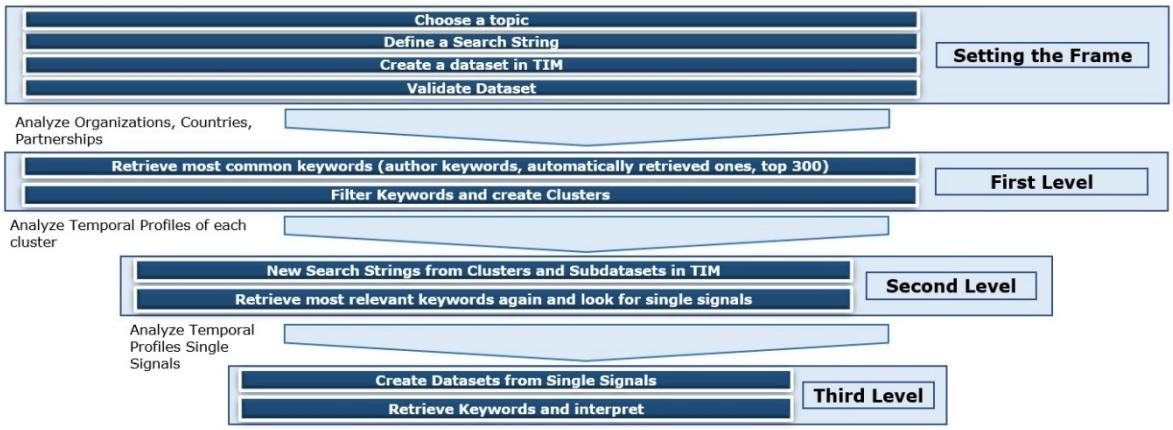


Figure 6: Designing a horizon scanning exercise in TIM.

After setting up and verifying the data set, TIM performs numerous analysis on the bibliometric data. Text mining techniques enable retrieving the relative frequency and importance of data components depending on sources (articles, conference proceedings, book chapters, reviews, patents, EU projects, etc.) or topics of interest, and their trends over time compared to others, identify the emergence of a topic or component within the data set.

Network analysis methods are relevant to give a synoptic view of the institutional landscape. Based on the author affiliations, network graphs are available on the organizational level as well as on the country level (and even Nomenclature of Territorial Units (NUTS) regions). On the organizational level, entities can also be filtered by different categories like university, company, hospital, research centre, foundation, and others. Network graphs are used to outline which organizations are driving research and innovation for a particular topic and how they are associated with each other in terms of partnerships.

One downside of using TIM for this exercise is the limited flexibility for data analysis. Though, even new indicators can be created in TIM and applied to the data set, other interesting analytical steps are missing like text similarity analysis to perform the clustering automatically and the availability of abstracts, and not only the retrieved keywords to implement it. In addition, the data set cannot be accessed through TIM, so underlying data of a publication is not visible. Thus, to circumvent this limitation, the Scopus API has been utilized in Python in order to create the exact same data set in Python, making it possible to access the meta data of each publication and retrieve the respective abstracts. Once these abstracts are made available, clustering and text similarity algorithms can be deployed to further automate the outlined workflow. Powerful text mining libraries like NLTK combined with data analysis packages like Pandas or Numpy and Machine Learning libraries like sklearn allow for endless possibilities to arrive at meaningful outputs that can be insightful for a horizon scanning (see chapter 6).

Due to time and resource constraints it was not possible to further test and explore these possibilities. Furthermore, a limitation of the Scopus API is that it does not allow to retrieve the full text of a publication, so the data input for the analysis includes at best, the title, keywords and abstracts only.

3. Horizon Scanning research and innovation on sustainable resources in Africa with an insight into the agricultural domain

3.1 Exploratory scanning process on sustainable resources in Africa

The horizon scanning exercise conducted in the context of the present report applied an exploratory scanning process that was framed around the overarching topic of sustainable resources with a geographic focus on Africa. The idea was to keep the topic as broad as possible to define a large data set in TIM from which weak signals can be identified. The workflow outlined in chapter 2.4 was applied and the search string of the data set initially addressed issues of water, soils, biodiversity, forests, land cover/land use resources, human activities like cities and transport, agriculture and aspects related to climate change, and marine resources. The dataset featured around 18.000 publication records from Scopus, CORDIS and PATSTAT, though largely dominated by publication records from the Scopus database.

After the dataset was created, keywords were retrieved and filtered. Due to the high number of publication records included in the dataset and the diversity of topics targeted, the retrieved keywords remained too broad and vague with a weak potential for capturing specific and sufficient detail and identify potential weak signals among them. This first outcome led to the adjustment of the process for the horizon scanning in a way that a funnel was applied to gradually increase the level of detail by using the filtered output from the retrieved keywords to create new datasets and repeat the process steps until weak signals were identified. However, this would imply a screening of several thousands of keywords for the entire topic of sustainable resources in Africa. For demonstration purposes, this funnel was applied to the topic of agriculture in order to evaluate the feasibility, usefulness, and efficiency of the approach. The figure below summarises the different steps of the scanning process with regard to the analysis of the content.

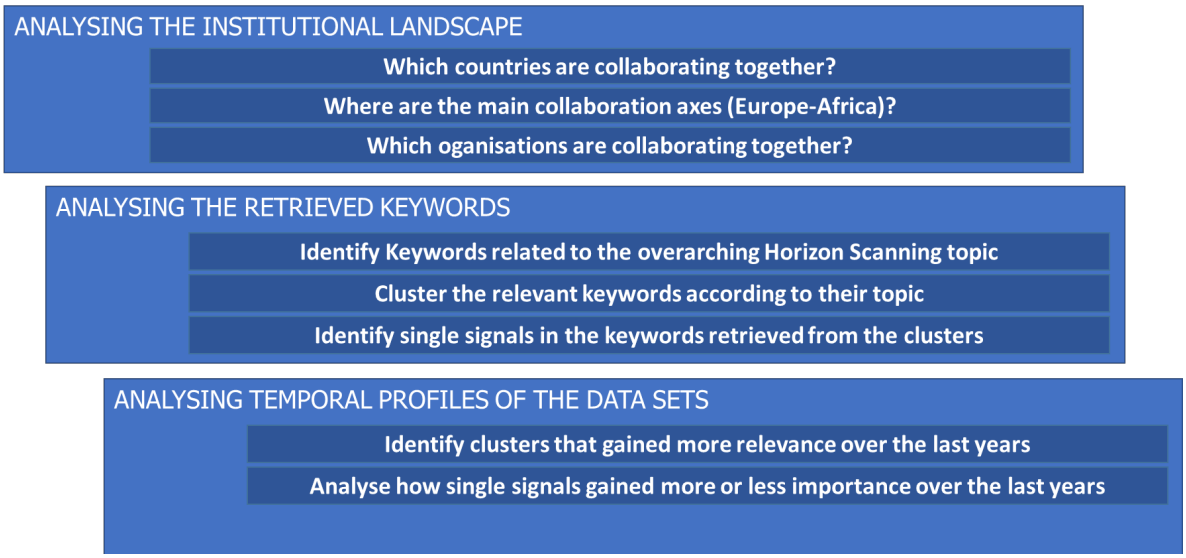


Figure 7: Content Analysis during horizon scanning exercise.

3.2 Case study on agricultural research in Africa

The workflow outlined in chapter 2.4 was applied to the topic of agricultural research and development in Africa and its outcomes are presented in this chapter. When the search string was designed, peculiarities of agriculture in Africa had to be considered. Though agriculture in Africa in itself is obviously quite diverse and relevant issues can differ greatly among different regions across the continent. The data set created in TIM from the search string described in section 2.4 featured 4493 publication records and keywords were retrieved automatically. Based on the number of occurrences across publications, aspects of land and tenure rights, land conflict, family or small-scale farming and food security, particularly nutrition and health usually played an important role in most places. It is important to note that the datasets were all created before February 2020 and their number of publications may vary by now due to updates of the data bases that could ultimately lead to potentially different results of the exercise as a whole. It will be interesting to re-create the datasets after summer 2020 to address the impact of the Covid-19 pandemic on agricultural research and related development in Africa.

3.2.1 Characterising the African research and innovation area

Back in the early 2000's (UNEP (2006)), experts reported that in Africa the lack of capacity, in terms of skills and opportunities, to manage environmental resources undermines the potential for sustainable development – consequently strengthening institutions and empowering people in development are important strategies for the future. It is acknowledged that in most African countries, there is a relatively low investment in research activities as a percentage of GDP. Today the situation in Africa remains difficult and low spending on research and development is still listed within the agricultural domain in addition to falling productivity, inefficiency of ongoing research in reaching the farmer; the need for reform towards sustainable research and its funding at national, sub-regional and regional levels; integrating technology adoption; strengthening institutions. The several lines of action in the African Union Development Agency (AUDA-NEPAD[3]) research programme to improve the situation includes the cross-cutting initiative on scientific capacity building. What is the diversity of the African institutional landscape now and can we sense its academic performance in the last five years? Recently, insight can be gained from the Third African Innovation Outlook that presents the status of Research and Experimental Development (R&D) and Innovation performance in Africa, focusing on the period from 2013 to 2016[4]. It has become crucial to improve knowledge and enhance capabilities through stable investments, the development of scientific partnerships and promoting a fair and equitable use of patenting systems.

What are the recent trends in these respects? Are institutions getting more proactive and building partnerships worldwide and particularly in Europe? Are there any impulse of new technologies and innovation systems, thus more patenting related to agricultural research? These are few questions that underpin the scanning exercise of the academic records, research projects and patents in Africa.

[3] https://www.nepad.org/publication/annual-report-2017

[4] https://www.nepad.org/publication/african-innovation-outlook-iii

Academic records, projects and patents

When looking at the distribution of different data sources within the data set, it becomes clear that Scopus features the majority of records, while EU projects from the CORDIS data base only provides a few records and patents are almost not present in the data set (see Figure 8).

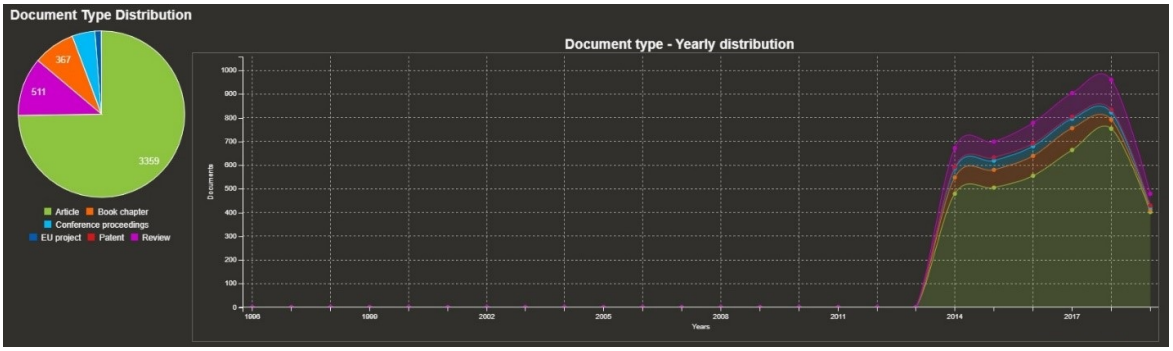


Figure 8: Document type distribution of dataset in TIM.

Generally speaking, the increase in patent filings in Africa can be observed in the graph below that shows yearly patent filings reported from all African countries since 1996 (dataset created in TIM about patent fillings in general, not specific to the agricultural sector) (see Figure 9). While the curve in the graph features an increase in the number of patents in the last five years, why the number of patents and share within the R&D sector remain so low? Are policymakers neglecting patent systems in Africa? This should be investigated given the increasingly important role that knowledge to innovate and create value plays in economic growth. According to Graff and Pardey (2019), only a small share of inventions globally is made in sub-Saharan Africa, but for those inventions that do arise in Africa, patent systems are weak and foreign filings are made widely outside of Africa in high income countries. Only in South Africa, between 15% and 20% of patent filings are by residents of South Africa, compared to 3% in other developing and emerging economies. Figure 9 confirms what Graff and Pardey (2019) have reported on the increase of patent filings in Africa, particularly since the mid-1990s and they added on inventions being mainly in biotechnologies for healthcare and bio-economy. 17% of biological inventions indicated an application in “Agriculture”, 6% in “food and beverage” and 4% in “environment”.

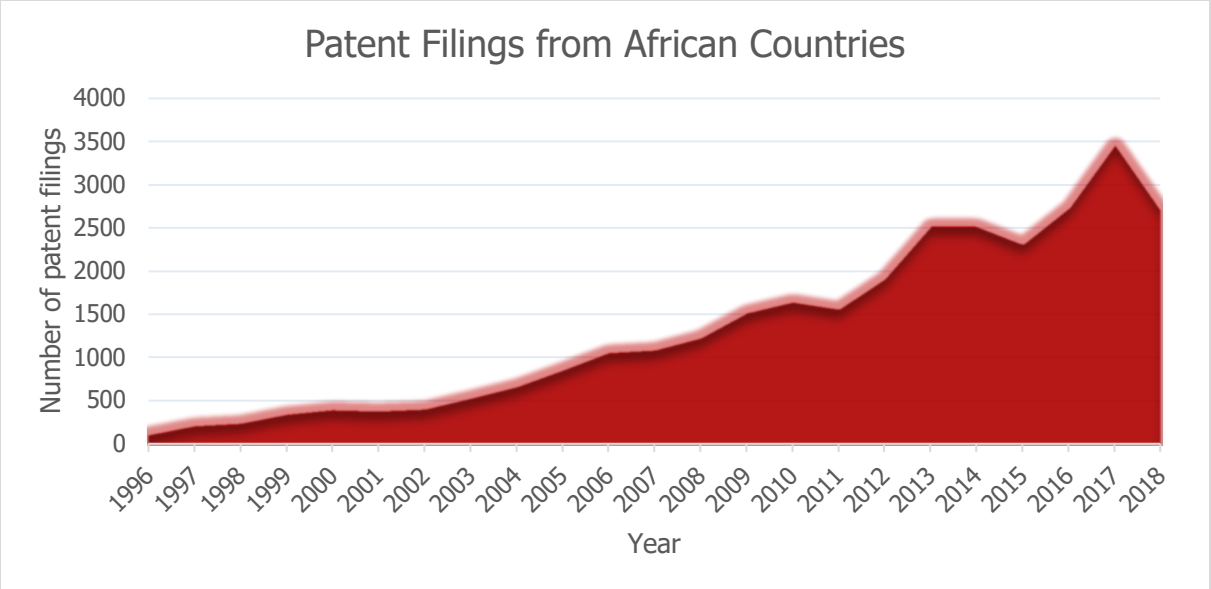


Figure 9: Patent Filings from African Countries based on TIM dataset.

The emergence and increasing numbers of publication records reflects an increase in research activities in general (see Figure 8). The number of universities have indeed increased worldwide, and every country now has a university, this is also true in Africa. The increase in records does not necessarily indicate that this particular topic (agriculture) became more relevant in the academic community over the last years. Therefore, it could be interesting to compare the rate of research activities with another topic. Further it would be relevant to compare the annual growth rate of publications with the GDP growth. Also the drop-in number of records observed at year 2019 is most likely irrelevant and explained by the 6 monthly updates of Scopus and by the time lag between the review and registration process of publications.
In addition, all publication records in the TIM based data set (reviews, articles, books, and conferences) deal with Africa but do not obviously include an author affiliated at an African University. Nonetheless, we can infer a correlation between the observed increase in records and the performance of academic staff, thus featuring improved staffing and related policies (recruitments, staff development, learning, international collaboration).

Institutional Landscape and Partnerships

The study continues with exploring the geographical distribution of the records to identify which countries conducted agricultural research in Africa and worldwide, where were the most proactive universities over the last 5 years and what were the main collaboration axes. The map below indicates the highest number of publications in traditionally prolific publishers of high-quality science worldwide like the European countries, Canada, the United States and Australia, but also in some emerging African countries. India, China, Japan, and South Korea also feature a slightly lower, still considerable number of publications, thus confirming the growing importance, role and interest of Asiatic countries for agricultural research in Africa.

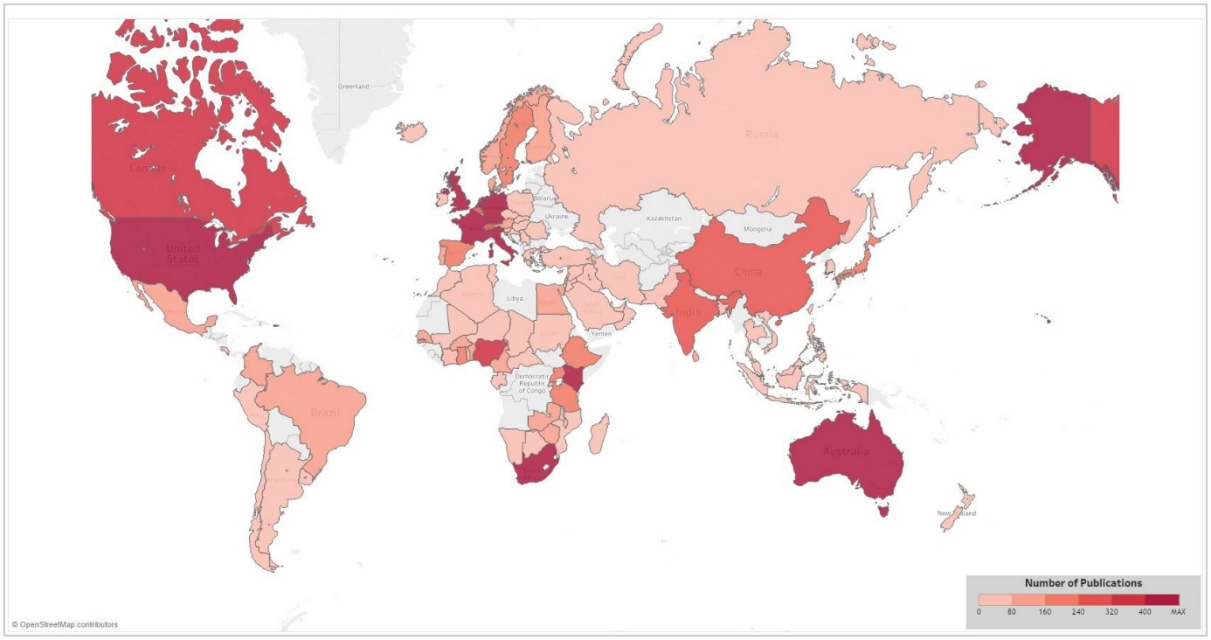


Figure 10: Distribution of publication records by country for dataset in TIM
(no data from countries in grey shade).

Looking only at the African countries, South Africa dominates the research on agriculture in Africa, followed by countries like Kenya, Nigeria, Uganda or Ghana (Figure 11). In figure 11, the larger size the circle is, the higher the capacity of the country to produce or collaborate on agricultural research and development.
Such diagram can help to identify and focus attention on the underperformance of individual countries, so that these countries may be targeted for additional assistance where it is most badly needed whether this be on research, technical innovation or financially. Best practices in high-performing countries may also be emulated by those countries showing weaker capacity and slower progress.
The level of economic development and investment in the research and development area explains the high number of publications in high-income non-African countries. In low-income African countries, it is not surprising to find countries with higher economic development allocating budgets to research and development.

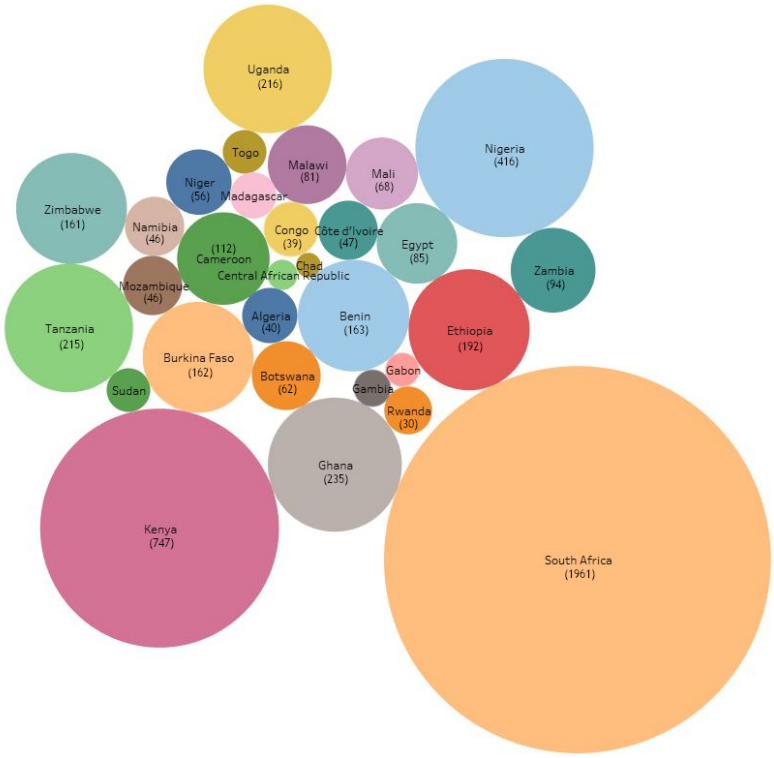


Figure 11: Number of publication records among African countries.

Our TIM based findings on African countries with higher research capacity are aligned with analysis from AUDA-NEPAD, 2019 where further insights are available. The dominance of South Africa in the agricultural R&D sector is explained by a stronger economy, but also by the comparably stronger R&D business sector where expenditures are relatively high compared to other countries with higher income levels like Egypt where the private sector R&D expenditures are fairly low. Likewise, R&D spending in the public sector is relatively high in South Africa (0.82% of GDP, second among African countries), but also in Kenya (0.98% of GDP, highest rate among African countries) that proves to provide a significant amount of agricultural research publications as well as shown in figure 11. However, insights apply to the R&D sector in general and do not specifically focus on agricultural research only. Furthermore, reliable data on R&D public and private spending is lacking in many other countries, as well as data on R&D personnel, making it thus, difficult to make respective conclusions as outlined in the graphs above (AUDA-NEPAD 2019). Nigeria's importance highlighted in figure 11 is confirmed as another important country driving agricultural research within Africa. Between 2000 and 2008, public R&D spending doubled, this is mainly due to an overall improvement of the economic situation in the country. The higher education sector appears to play an essential role in the field of agricultural research in Nigeria and the personnel employed in the R&D sector also increased continuously. R&D investments are mainly carried out by the government, while business sector engagement is lagging behind.

Overall, R&D spending in African countries largely depends on economic and political stability and ruling governments largely determine R&D activities in the countries (Flaherty et al. 2010).

In a next step, the distribution of publication records was broken down to the institutional level. The top 20 organizations bring a different perspective, showing that mainly African universities and research institutes dominate the research activities in this field (see figure

12). This indicates that apart from Wageningen University or the University of California which still feature relatively high numbers of publications, no actual research hotspots on agriculture in Africa exist outside Africa and the high numbers of publication records in countries like France, Germany or the UK are most likely more evenly distributed across various institutes and universities with fairly low numbers of publications. Surprisingly, the JRC is not listed among the top 20 organisations in the dataset.

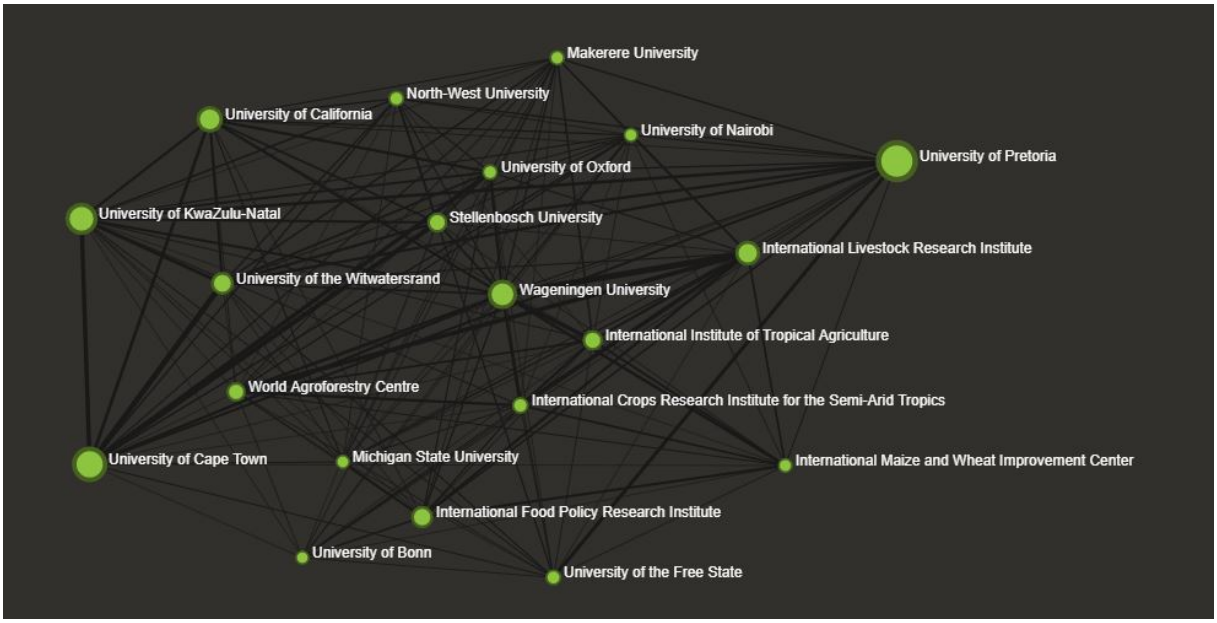


Figure 12: Network Visualization of Top 20 Organisations according to their number of publications for the dataset in TIM.

Summarising the findings derived from the graphs above, there is a strong contribution from Western countries in the field of agricultural research in Africa, with Europe being one of the main collaborators, followed by the United States. Inside Africa, South Africa is playing a dominant role within agricultural research, partially due to its economic development, but also due to higher private sector engagement in R&D activities compared to other countries. Nonetheless, countries like Nigeria and Kenya are also important contributors in the field with Kenya allocating even more of its budget to research activities in relation to the GDP. This constellation is also reflected on the institutional level, where South African universities feature the highest number of collaborations with other universities, though also research facilities like the International Livestock Research Institute in Kenya or the International Institute of Tropic Agriculture in Nigeria show a higher degree of connectedness with other institutions. For the collaboration axis between African and European organisations, Wageningen University plays a central role which makes sense, given the fact that this is meant to be one of the best agricultural universities in the world. For the collaboration axis between American and African institutions, the University of California is an important collaborator. All in all, conclusions on political partnerships between countries or organizations that are derived from bibliometric data always need to be drawn carefully. Further, when more data will be available, a dedicated TIM exercise to the solely African records would be also relevant to better feature progress in the collaboration between African countries only. To conclude, the currently available data prove to support the horizon scanning exercise in further contextualising the topic and to contribute to a better understanding of relevant stakeholders and their constellations and collaborations.

3.2.2 Keyword Clustering

Africa holds more than 60% of the world's arable land but its land productivity remains low. According to experts, modest yields result, to a large extent, from the low use of fertilisers, a lack of mechanisation and irrigation, and the impact of climate change. Underinvestment in the sector and poor governance are additional explanatory factors despite around two-thirds of the African population is employed within this sector, the vast majority working on small-scale farms.

According to the "Agriculture Outlook 2018-27" report from the OECD and the UN Food and Agriculture Organisation (OECD/FAO (2018)), the sector will undergo robust growth, with crop production in sub-Saharan Africa projected to rise 30% between 2018 and 2028, and innovation uptake in smart farming techniques.

However, with the population expected to double by 2050 reaching over 2 billion people, and with climate change anticipated to have deleterious impact, crop production - mostly from small farm holders - will hardly meet domestic consumption demand that will continue to depend on global markets and imports over the medium term. African governments will need to increase investment in agriculture, including in infrastructure, while also supporting the development of agri-business. These changes could help the agricultural sector to play a truly transformative role in supporting Africa's successful long-term economic development.

Within the Third Africa Environment Outlook from the United Nations Environment Programme (2013), experts had further analysed the importance of, and interlinkages between, health and environment and the opportunities and synergies that might be derived from intensified collaboration between the two sectors. More recently, the Natural Resources Governance and Food Security thematic area under the Comprehensive Africa Agriculture Development Programme (CAADP)[5] has set goals and targets to increase by year 2025 the public investment in agriculture by a minimum of 10% of national budgets, and to raise agricultural productivity by at least 6%; their biannual progress reports of 2017 and 2019 includes Country Scorecard for implementing the Malabo Declaration on Agriculture transformation in Africa[6].

The actual horizon scanning started with the idea to identify the main issues at stake and challenges ahead within the agricultural domain and any emerging signals of the rural transformative changes reflected in the research domain. The 600 keywords - the top 300 keywords provided by the authors and the top 300 keywords based on the TF – IDF scores of the terms were screened and evaluated. The default settings in TIM usually allow the user to only retrieve the top 100 keywords, but by extending the user capacities, the top 300 keywords were retrieved.

This was very useful for the horizon scanning exercise, since potentially interesting signals are not necessarily obtained from the most important keywords from the corpus but can be more concealed at the bottom of the keyword lists as well.

After screening through the keywords, filtering them and eventually clustering them, the following thematic clusters were identified (figure 13):

[5] https://www.nepad.org/caadp/overview

[6] https://www.nepad.org/caadp/publication/2019-africa-agriculture-transformation-scorecard-aats

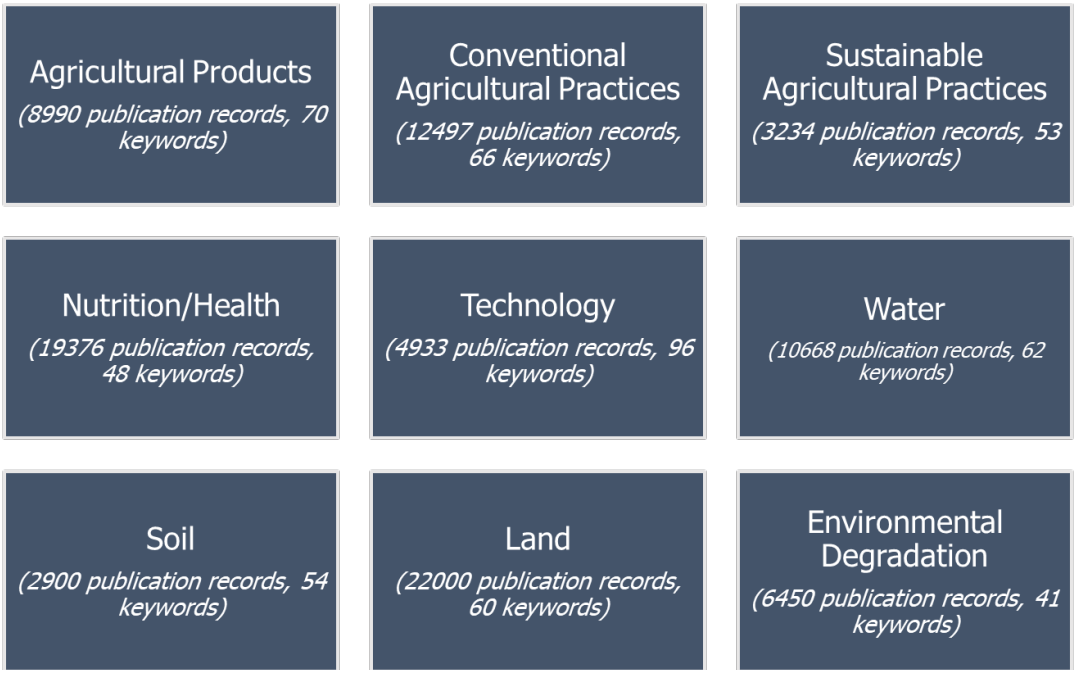


Figure 13: Thematic clusters characterising agricultural research in Africa after keywords screening.

Agricultural Products

(8990 Publication Records, 70 Keywords identified)

The keywords for this cluster featured two broad thematic research directions: one on the products themselves and one on the threats associated with these products. **Maize** and **livestock**, along with **Cassava** were brought up most prominently, but also other products like **cow peas, pearl millets** or **cereal crops** showed up in the keywords of this cluster. The majority of keywords was noticeably focused on bacteria, insects and other viruses and pests threatening either crops or livestock. Especially **ticks** and **mosquitos** (also **tsetse fly**) as transmitters of those diseases were highlighted by the keywords. Example of diseases mentioned were **cassava mosaic disease, cassava brown steak disease, east coast fever, arbovirus, African swine fever** or **lumpy skin disease**; regarding pathogens **trypanosoma congolense** was often mentioned and other example for ticks **(rhipicephalus microplus**), bacteria (**rickettsia)** and parasites (**theileria parva)** were provided as well. As a matter of fact, biological and epidemiological concerns are identified as a priority field for research - in particular the increasing emergence or re-emergence of infectious diseases in the human population since the beginning of the 21st century in Africa , 60% of which are zoonoses (Fenollar and Mediannikov, 2018), the increased importance and frequency of livestock diseases at the wildlife-livestock interface (Swei et al, 2019), and the increased innovations in this field are particularly acknowledged in Graff and Pardey (2019). Last but not least, some keywords also focused on aspects of responding to these threats (e.g. **Integrated Pest Management, antimicrobial resistance**).

Conventional agricultural Practices

(12497 Publication Records, 66 keywords identified)

The keywords in this cluster give an idea of the structure of the conventional agricultural activities in Africa, ranging from **pastoralism, subsistence** and **smallholder farming** with **crop systems** where **conventional tillage** is applied.

Additional agricultural practices referred to **aquaculture** (the farming of aquatic organisms e.g. **fisheries and of aquatic plants** and **horticulture**. Furthermore, **foraging** is featured numerous times in the dataset which means it is a relevant activity not only for the livelihoods of hunters and gatherers but also for contributing to food security in general. There are many innovative efforts made today to cope with the challenges people in Africa are facing as confirmed in Hitchcock (2019). Overall, an increasing demand for agricultural products indicates a guiding paradigm of **agricultural intensification** that focuses on improvement of **seeding and crop systems** in order to increase the **agricultural productivity** (for instance by using **inorganic fertilizers** and **pesticides**). Aspects of **biological invasions** and **infestations** were also highlighted within the keywords, along with issues of **domestication**. Finally, keywords also referred to the importance of **agricultural investments** and **agricultural research**.

Sustainable Agricultural Practices

(3234 Publication Records, 53 keywords identified)

As expected, aspects of environmental degradation and impacts of climate change on agriculture were brought up massively and keywords were numerous to illustrate alternative agricultural practices meant to be sustainable. In this context, the concepts of **conservation agriculture, sustainable intensification, climate smart agriculture, agro-ecological zoning, sustainable land management, urban ecology, payment for ecosystem services, crop diversification** and **rotation** as well as **intercropping, new green revolution, bio fortification** and **agro-forestry** were mentioned. For managing the crops, aspects of **crop modelling** were addressed as well as the application of **minimum** or even **zero tillage** and **crop residue retention**. As mentioned before, for dealing with threats to the crops, yield or livestock, **integrated pest management** or the **sterile insect technique** came out as suggestions from the keywords. Another thematic streamline in this cluster dealt with aspects of **agricultural modernization, agricultural technology, agricultural innovation platforms** and **agricultural transformation**[7]. The current trend on innovative sustainable agriculture and insight on nature-based solutions are confirmed among others in Sonnenschein, L., Tiberious Brian, E. 2019.

Food nutrition and health

(19376 Publication Records, 48 keywords identified)

Food security is made of four components: (1) food availability including production, distribution and exchange, (2) food access including affordability, allocation and preference (3) food utilisation with nutrition, health and food safety (4) stability. Most records in this cluster are focused on food utilisation, particularly nutrition and health related issues. **Malnutrition, stunting** and **hidden hunger** were raised, as well as food-related health problems or problems related to diets **like obesity, vitamin a** and **d deficiencies**, **hypertension** and **diabetes,** altogether highlighting aspects of **nutrition transitions**. Underlying causes driving food insecurity and related to food availability were also detected in relation with other clusters, like soil degradation or agricultural intensification.

[7] http://www.fao.org/fao-stories/article/en/c/1170362/

Technology

(4933 Publication Records, 96 keywords identified)

With regard to crop modelling, most prominently, aspects of earth observation (**remote sensing, geospatial analysis)** were addressed, mentioning some of the respective satellites or their instruments/sensors being used for the data acquisition (**sentinel 1, 2, landsat 8, modis, avhrr, lidar**). For the analysis of the data, **trend analysis, change detection** (especially **NDVI, leaf area index, enhanced vegetation index**), **image classification, spectral mixture analysis** and **spatiotemporal analysis** were brought up. As monitoring instrument for the environment, **wireless sensor networks** were mentioned and with further regard to monitoring aspects, **drought early warning systems** also showed up in the keywords. **Big data** and the issue of **open source** data and software was also addressed in this context as well as the importance of **ICT4D**. Furthermore, the importance of **mobile phones** (relevance of **SMS**) for accessing data and related information was highlighted, as well as aspects of **mobile learning**. **E-Health, health informatics, telemedicine** and **E-governance** were also raised in the keywords, though the search string did not specifically look for health related issues (apart from food security in general). Additionally, crop models like **DSSAT** (Decision Support System for Agrotechnology Transfer) that aim for improving agricultural crop growth and estimating the efficiency of different farming methods were featured among the keywords.

As a matter of fact, mobile phone technology massively spread across the African continent, enabling an increasing number of people to get access to mobile phones. This access had several social and economic impacts on people´s lives, ranging from health improvements due to health information campaigns, impacts on personal safety (access to information on natural disasters), improvements in democratic structures (election monitoring) to a facilitated access to finances (mobile banking) and education (mobile learning) – overall enhancing interconnectedness and the potential for information and communication (Manacorda & Tesei 2020).

Aspects of telemedicine offer a great potential to provide access to scarce medical specialist care, especially in rural African areas and to counteract the notorious shortage of doctors in these regions. Nonetheless, numerous challenges related to legal and ethical concerns remain and the uptake of this new approach has been limited so far (Mars 2013). On the contrary, mobile learning gained more and more ground in the recent years and became an integral part of educational systems across African countries (Kaliisa & Picard 2017).

It is interesting to note that three emerging technologies prioritised for development in the Industrialization, Science, Technology and Innovation programme of the New Partnership for Africa's Development (NEPAD) (nepad.org) were not identified within the keywords characterising the research area; they are drones for agricultural transformation particularly for precision farming (https://www.nepad.org/publication/drones-horizon-transforming-africas-agriculture), gene drive to control and eliminate malaria (https://www.nepad.org/publication/gene-drives-malaria-control-and-elimination-africa-0 and functional biosafety systems established and supported in 17 countries more than 10,000 biosafety regulators trained on the application of agricultural biotechnology according to the progress (https://www.nepad.org/areas-of-work/industrialisation-science-technology-and-innovation).

Further, keywords did not cover technologies and innovations related to "bio-economy" concepts while there is a great potential and associated technical challenges for food and non-food biomass production and processing in Africa (Callo-Concha et al, 2020).

Water
(10668 Publication Records, 62 keywords identified)
This cluster captures four aspects associated with water resources: water resource management (**groundwater management, integrated water resource management, irrigation, water use efficiency, water footprint, water harvesting, damming, hydropower**), water availability (**water stress, water scarcity, droughts, drought tolerance, drought stress, precipitation, rainfall variability, seasonality, extreme precipitation, flooding, tropical rainfall measuring mission**), **water governance (water policy, hydro politics, transboundary rivers)** and **wastewater** management (**urban water, acid mine drainage**). The latter does not point solely at sanitation purposes but also refers to water reuse as a research theme to underpin urban-rural linkages, as well as at the removal of metals in water use for irrigation of agricultural areas.
De facto, problems of water resource management in relation to water use efficiency, ground water management and water availability become particularly visible in North Africa that is the driest region of the continent with very limited water resources, which amplifies the region's vulnerability due to climate change related impacts and risks (Chaibi 2018). Extending the perspective of water resource management to the institutional level, water governance is confirmed as a crucial issue to address. It becomes a politically sensitive process in which power asymmetries and a lack of institutional coordination on various levels hamper effective water management in drought-haunted areas like rural South-Africa (Förster et al. 2017, Makaya et al. 2020).

Soil
(2900 Publication records, 54 keywords identified)
In the soil cluster, the soil's capacity of serving as a **carbon sink** was highlighted (**carbon sequestration, soil organic carbon, carbon stock**). Secondly, the role of soils in different material cycles was addressed (**phosphorus cycle, nitrogen cycle, carbon cycle, nutrient cycling, soil microbial community**). The majority of keywords dealt with the quality of the soil **(soil moisture, soil fertility, soil erosion, soil degradation, soil restoration, sandy soils, integrated soil fertility management**) without however mentioning soil biodiversity. Last but not least, the use of fertilizers and their impact on the soil was highlighted as well (**farmyard manure, nitric oxide, inorganic fertilizers)**.
Indeed, the interaction between agricultural practices, nutrient cycles and carbon sequestration inside the soil is highly complex and can have different impacts on soil degradation and crop productivity, depending on the deployed practices that need to be tackled (Nafi et al. 2020). Risks of soil degradation and loss in soil diversity and productivity are particularly present in areas that feature relatively low soil nutrients, organic carbon and clay contents in general. Paired with growing populations and increasing food demands, the fertility of cultivated soils and its associated land resources is under high pressure, ultimately jeopardising food security of the respective population, as can be observed in many countries across Sub-Saharan Africa (Bado & Bationo 2018).

Land
(22000 publication records, 60 keywords identified)

In the land cluster, the question of ownership was addressed through numerous keywords (**tenure rights, land grab, land reforms, land acquisition**). Aspects of **rural development (rural electrification, spatial planning, rural – urban, migration)** were highlighted as well. Urbanization showed up as another pressing topic in the African context (**rapid urbanization, resettlements, dispersion, urban expansion, slums, informal settlements, urban informality, townships, urban ecology, urban governance**), particularly emphasizing the need for land governance and concerns at the rural-urban fringe. Combined with population growth, this leads to **land use conflicts** and competitions with agricultural areas.
The question of ownership is confirmed as a hot topic in Africa that causes tensions within societies and political systems. Furthermore, Western concepts of tenure rights and ownership cannot be equally applied within the African context and terms like tenure, freehold, usufruct and ownership are still characterized by discordance when they are applied in practice (Davison 2018).
De facto, the rural-urban interface is dominated by rapid urbanization that is fuelled by rural-urban migration processes that reversely shape rural transformation processes in the areas of origin. Urbanization in developing countries is the primary driver of land use changes and agricultural land conversion, especially in the peri-urban areas where the dynamics in land use changes are occurring at a rapid rate, potentially leading to air pollution, biodiversity loss, soil degradation and a deteriorating water quality (Abo-El-Wafa et al. 2017). On the other hand, there are increasing efforts to boost rural development, for instance by improving rural electrification.
Rural electrification has been identified as a key enabler for rural communities to foster sustainable development that can positively impact communities' health, education, and economic activities (Kyriakarakos et al. 2020) and research is needed as refer to in Moner Girona et al, 2018: *..there is a need to accelerate the pace of rural electrification in order to achieve full access to electricity by 2030, especially rural areas in Sub-Saharan Africa that has the least electrification rate globally. Photovoltaic hybrid mini-grids for example offer a modular and competitive solution to accelerate the electrification in rural areas, characterised by remoteness and sparse population density. Research is still needed to serve multipurpose to inform investors, energy planners and policymakers.*

Environmental Degradation
(6450 Publication records, 41 keywords identified)
The last cluster deals with the environmental impacts resulting from the agricultural activities. One thematic stream addresses the loss of natural resources and environmental degradation (**soil degradation, deforestation, desertification, habitat loss**) and their drivers (**agriculture, urbanization, mining, wood fuel, bio fuel, cooking stoves, hydraulic fracturing**). Another important topic concerns pollution and waste (**waste management, e-waste, solid waste, wastewater, waste to energy, mine wastes, composting, contamination**) directly or indirectly impacting agricultural lands.
As a matter of fact, the interplay between agriculture and environmental degradation has been subject to numerous studies. Olanipekun et al. (2019) analysed the connection between agricultural practices, poverty and environmental degradation in 11 Central and West African countries and found that rising income levels can reduce the negative impacts of agriculture on the environment. Furthermore, increased use of renewable energy resources and adequate regulations also dampen environmental degradation. Conversely,

population growth drives environmental degradation. Poverty alleviation and mitigation can contribute to avoid environmental degradation while unsustainable agricultural practices in general aggravate negative impacts on the environment (Olanipekun et al. 2019).
Waste is another area of active research, particularly relevant in the context of emerging African bio-economies (Callo-Concha et al, 2020) although as previously said, the keyword "bio-economy' did not show-up as one solution to the problem. Waste prevention as well as its efficient management as input for renewable bio-based products with close to zero greenhouse gas emissions are major challenges in Africa and in the world in general. The increase of waste production, including food loss and waste as well as the mismanagement of both organic and solid waste are a reality in African rural and urban environments leading to environmental and health impacts like arbovirosis epidemics due to proliferation of mosquitoes, water and soil pollution and methane emissions due to the degradation of waste in open dumps (Le Picard, 2019). A more efficient waste management including collection, selective waste sorting, treatment, and recycling as well as the development of waste-to-energy plants to generate electricity are relevant solutions in the context of biomass-based economy. Modern biomass processing is still in an early stage in Sub-Saharan Africa and most related matters are rooted in the agricultural sector.

3.2.3 Single Signals

In a next step, sub data sets based on this clustering output were created and the process of keyword retrieval, screening and filtering repeated. The final output of this process led to the identification of isolated keyworks with few records (in the range of 1-10 publications) that were considered as weak signals potentially providing useful insights for the horizon scanning exercise. Some of these weak signals included:

Figure 14: Weak Signals detected from the clusters.

Applying the funnel approach (third level of Figure 6), a single signal is then being selected and a respective sub data set created in TIM in order to repeat the process carried out in the two steps before. The final list of obtained keywords provides a comprehending context for the respective signal. Taking **Rural Transformation** as an example, obtained keywords for this term provide more details on pressing issues for this signal like aspects of youth unemployment, the rural-urban interface, education, entrepreneurship development and migration governance.

As we modernize food systems to make them climate-smart, healthy, and sustainable, we must also strive to make them inclusive of smallholders, youth, women, conflict-affected people, and other poor and marginalized people. This statement taken from the April 2020 synopsis of the International Food Policy Research Institute (IFPRI)[8] identify issues like inclusive rural transformation, as well as gender balance and combat inequality as crucial for the future.

Green Manure was another single signal identified from the cluster keywords. The keywords related to Green Manure focused on soil properties and soil management practices that consider the sustainable usage of fertilizers (biological nitrogen fixation, microbial activity, residue retention, mineral fertilizers) to avoid soil erosion and soil disturbances and improve soil fertility. The International Center for Tropical Agriculture (CIAT) notes that smallholder farmers in Sub-Saharan Africa usually have limited access to mineral fertilizers due to lack of financial resources, whereas in other examples the inadequate use of mineral fertilizers combined with unsustainable agricultural intensification has led to increased soil infertility. *Green manure/cover crops (GMCCs) like Lablab bean, pencil flower or velvet bean offer a great potential for contributing to soil health and food security due to their soil amelioration effects and affordability (CIAT 2019).*

As a third example, **Nutrition Transition** highlights shifts in diets that affect food systems on a large scale. Keywords related to this transition address industrial agriculture and meat production that causes the loss of traditional food and feeding practices as well as dietary guidelines and eating patterns with regard to obesity and overweight. On the contrary, globalization has also positively impacted the reviving of traditional food by emphasizing local champions and mobilizing international support to revive indigenous knowledge and specific technologies to strengthen local farmer's capacities contributing to local food systems (Trolio et al. 2016). In the field of Nutrition Transition, strategic research is developing in the context of (re) shaping food systems as stated by the International Food Policy Research Institute [9]*: Food systems must be reshaped to provide access to nutritious, affordable, and safe foods to enable all individuals to achieve their full potential as healthy, productive members of society and to ensure the health and nutrition of future generations*.

The example of **Blue Carbon** provides another interesting signal, that focuses specifically on mangrove ecosystems and their agricultural potential (shrimp farming) as well as their carbon storage potential. The keywords associated to Blue Carbon also indicated that the ecosystems are endangered by deforestation activities and efficient monitoring through earth observation data and methods (Terrasar X, Geoscience Laser Altimeter System, TanDEM-X Mission) are relevant (and necessary due to difficult access) to protect these coastal wetland ecosystems from being overexploited (Vasconcelos et al. 2015, Ahmed & Glaser 2016).

[8] https://www.gfpr.ifpri.info

[9] https://www.ifpri.org/strategic-research-area/promoting-healthy-diets-and-nutrition-for-all

The approach outlined above, was applied to all identified signals to characterise each of them in more detail and thus contributing to shape the outcome of the overarching horizon scanning exercise. Nonetheless, such results cannot be considered standalone and should rather serve as preliminary alerts and inputs for stimulating discussions among experts and decision makers, to raise awareness on underlying issues and eventually translate them into actionable knowledge for elaborating strategies and policy recommendations.
Furthermore, there were a few isolated keywords that could not be assigned to a cluster, nor could be considered as a single signal. Still, they remain important challenges that should not be overlooked when targeting the social process within the agricultural domain. They include aspects of gender and inequality, colonialism, participatory approaches, indigenous knowledge, agricultural knowledge gaps, as well as vulnerability and resilience of the poor, social capital, and capacity building. The gender and inequality issue - in particular integrating women small holder farmers into the mainstream of economy and the whole agricultural value chain as well as strengthened capacity of gender in ministries - is acknowledged as thematic goal for year 2025 in the "gender climate change agricultural support" programme (see https://www.nepad.org/programme/gender-climate-change-agriculture-support).
Additionally, the keywords **Maasai** and **Fulani** showed up, most likely to point at the nomadic pastoralism of these ethnic groups in drylands in Africa; it is a sensitive issue of land conflicts and unsustainable cattle management pointing at the need of integration between grazing livestock, crops and wildlife management.

3.2.4 Innovation Uptake

Another goal when conducting a horizon scanning exercise is to assess the uptake of innovation inside particular topics. For this aim, an indicator has been formulated as the simple ratio of the number of publication records obtained with a search block on innovation/novelty divided by the ones without this search block.
This straightforward approach allows for comparison between the topics. The results showed that clusters like "Sustainable Agricultural Practices" or "Environmental Degradation" feature a higher share of innovation in the data set than clusters like "Conventional Agricultural Practices" or "Agricultural Products" (see figure 15).
These trends are partially reflected within agricultural development agendas in Africa that increasingly recognize the impacts of conventional agricultural practices on environmental degradation and thus, call for a shift towards advanced technologies refocusing on agro-ecological approaches and Green Revolution technologies (Adenle et al. 2019).
However, agricultural investments in R&D in Africa relative to a country's GDP still remains significantly low. Due to time lags between the moment investments are being made and resulting impacts that can be observed, those investments are often made too late and only to a minor degree. When looking at Sub-Saharan Africa, recent data on agricultural research investments even indicates a slight decline in investment rates since 2014 (Africa Agriculture Status Report 2018). This lack of investment is also reflected in a lack of public expenditure in the agricultural sector in general, which is contradicting the intended rate of 10% laid out in the CAADP (Shimeles et al. 2018).

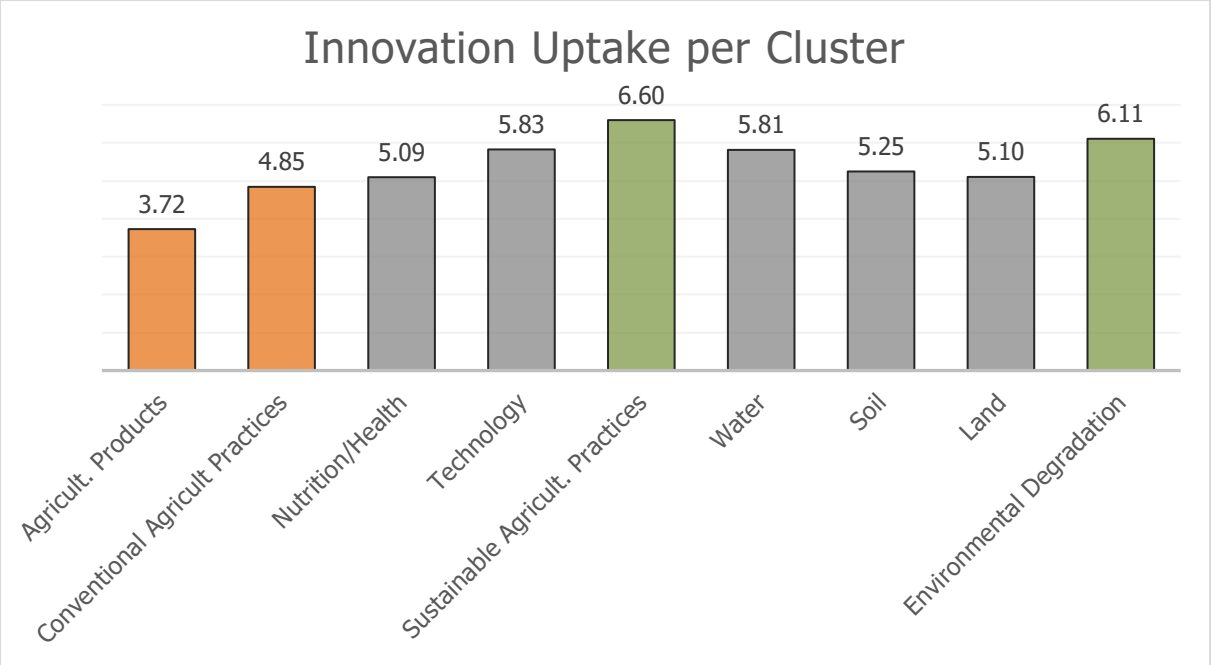


Figure 15: Innovation Uptake of clusters identified in TIM (higher rate higher uptake)

The applicability and reliability of this indicator now needs to be tested in different contexts; its ability to capture and compare the uptake of innovation across topics and time periods, is to be assessed in order to strengthen its usefulness in general and its capacity to monitor progress over time.

4. Benefits and limitations of Text Mining for Horizon Scanning

The applied approach brought up some issues and potential downsides that will be discussed in this chapter from both the methodological and thematic perspectives. The positive aspects and the benefits of the approach will also be reported.

Search string driving results

One major point that was raised earlier and that calls for caution is the construction of the search string. As the process is designed like a funnel it means that constructing the search string as an initial step of the process will trickle down through each of the following steps and ultimately drive the final results. Inevitably, specific terms used in the search string of the first round were also reflected in the obtained keywords. However, the funnel approach gradually detaches the final results from the original search string as new keywords are obtained after each round that serve as new input for creating the search string for the next round. This is acknowledged in previous studies which analysed the first output generated by the first search string in TIM and found sufficient results within the retrieved keywords (Moro et al. 2018). Alternatives exist to address the issue of the search string construction and will be discussed in chapter 6.

Data Sources utilized in TIM

The three different data sources used in TIM are very comprehensive and particularly useful in the context of horizon scanning but there are some downsides due to the fact that they do not provide the full picture. The data sources vary greatly in size which is why the Scopus data base with its exhaustive academic publications usually has a higher weight within the created data sets in comparison to projects and patents. The data is provided with considerable time lags that can interfere with the scope of a horizon scanning activity. The five years' time period selected in the present exercise may be too short to capture trends for this reason, in addition it only included one European Framework Programme (Horizon 2020). Alternative approaches based on expert workshops and group discussions solve partially the timeliness of information by conveying the scientific expertise and knowledge of experts in real time, nevertheless apart for their own research activities, experts rely on the same scientific publications than the ones available in TIM.

A data source that is potentially valuable for horizon scanning is the field of grey literature in which outcomes of international organisations, institutions, NGOs etc. are captured. The output is usually not standardised or peer-reviewed and the way these outputs are being disseminated and distributed can vary greatly, making it hence, a largely unstructured data source that cannot be accessed through a single platform, portal or whatsoever. Another data source that can be useful is news data from media coverage. A great advantage is the near real time in which the data is generated and disseminated. However, the timeliness compromises the reliability of the data and with recent problems related to the growing emergence of fake news being spread across the web, utilising this data source for horizon scanning can become complicated.

Flexibility in TIM

TIM is an easy-to-use tool to analyse bibliometric data in an efficient way particularly suited for people without particular backgrounds in data analysis and text mining. User capacities can even be extended, allowing to explore data in detail and to create individual indicators. On the other hand, all the calculations are performed on the data in the backend preventing the user from exploring the whole text of each publication record.

An open access to the underlying data would allow users to design their individual text mining workflows and adapt algorithms and calculations in accordance with their needs and scopes. Though these restrictions obviously arise due to copyright issues, an API could potentially allow users to use the data for their purposes without directly accessing them, by letting them apply their algorithms and obtaining the results at the end.

The human factor

Several steps within the workflow of the presented horizon scanning exercise are automated, nevertheless few essential steps where interpretative reasoning and complex assessments are needed require the guidance of experts. This concerns the construction of the search string, as well as the analysis of the keyword lists and the ultimate interpretation and sense making of the final results. There is thus a potential bias that is introduced by the experts and analysts.

Different scope, different method?

One of the most important insights from this exercise is the need to finetune the methodology according to the broadness and specificity of the topic limitations that may result from the very nature of the selected topic itself and the scope of the horizon scanning. Moro et al. 2018 explored emerging photovoltaic technologies by analysing bibliometric data with the help of TIM and managed to obtain sufficient and timely results, particularly emerging signals that were similar compared to the outcomes of an expert review. The focus on one specific technology (also being the innovation in itself) reduced the number of needed iterations and one iteration proved sufficient. In the present example, the topic of agriculture is in itself highly complex and comprises various sub-categories as indicated by the identified clusters. Furthermore, limiting the topic to a geographic extent adds complexity and context to the topic. For these reasons, a funnel approach seemed to be most suitable for tackling broad and multi-faceted issues and for contextualizing obtained keywords in a comprehensive way to gain more insights on the bigger picture. More horizon scanning exercises need to be conducted in different thematic and geographic contexts to support this conclusion. For example, it would be interesting to analyse the institutions' network within CORDIS data only. The research grants from the EU may reveal to be concentrated in a limited network of institutions, and the TIM exercise may help providing guidance on where to support the development of stronger research capacities in Africa. Targeting innovation is also discussed in the next section.

Getting novel thematic insights or pulling out old chestnuts?

Could the results be of value as input to reviews, debates, and assessment of activities for the preparation of initiatives in the targeted research domain (i.e. agriculture and associated cross-sectors)?

First, the scanning analysis on African research and innovation provides a valuable synoptic insight on the current institutional landscapes and networks, particularly in terms of most participative countries, best performing institutions, main active partnerships, and collaboration axes. Policy makers, such as the European Commission' s Directorate-General for Research and Innovation, could use this information to design research programmes that use in an efficient manner the existing networks and identify weak connections in need to be strengthened. Once identified, countries acknowledging weak participation could be targeted for new research calls. Improved understanding of the structure of these networks can help to boost research capacities within Africa and not only within consolidated networks that already exist.

The exercise could help promoting new partnerships and collaboration axes for the sharing and exchange of data and information, to better disseminate high quality science by facilitating the undertaking of co-operative national and international scientific research programmes. By identifying and showcasing where research has been successfully conducted, it could emulate and stimulate best practices in poor-performing countries. Better insight on innovation including patenting would require access to usually more confidential data from the business and industrial sectors to grab a complete and more reliable picture.

Second, the thematic clustering finds its value added in providing a solid synoptic view on the overall structure and main subthemes in the last 5 years (2014-2019) of the agricultural sector in Africa. It mainly confirms, consolidates, and updates knowledge about thematic matters; mostly were long-standing well-known matters, with some urging matters but always disregarded. We noticed that few known issues were missed out and not translated by keywords (like bio-economy, or components of food security other than nutrition and food production). The relative weight of clusters was surprising such as for example the cluster sustainable agricultural practices versus the conventional ones.

When running the exercise and obtaining keywords for each cluster, the presence of the same few keywords within each cluster - namely specific country names, climate change, HIV, malaria, non-communicable diseases - was noticeable, thus indicating the horizontal importance of those keywords for the African context in general. These keywords do not feature any new trends or upcoming challenges but rather illustrate consolidated or persistent, still unresolved, challenges for Africa specifically and worldwide in general. Furthermore, the presentation of the thematic clusters also showed the difficulties that potentially arise when analysing them as all of them are multi-dimensional in their nature and feature a high degree of complexity. Experts with specific expertise in each of the clusters could greatly contribute to the interpretation and contextualisation of these results. By identifying relevant and emerging fields of research, TIM provides a convenient way based on which further analysis and interpretation can produce valuable insights that can feed into policy discussions where remedial policy or management action can be anticipated or adapted. New technologies and innovation research approaches can be detected, and their application promoted accordingly. Overall, guidance on development projects, management decisions and policies can be provided if experts utilise the tool and its outcomes in the right way.

However, obtained results need to be treated with caution as the analysis may entail a potential caveat: topics dominating in terms of frequency may only indicate a topic that used to be hot in the past due to the timeframe of publishing the information and may already loose relevance in research by the moment it gains momentum in the literature. To better

catch up on potentially innovative issues in which the agricultural sector in Africa could shift to, it can be useful to put the focal point of analysis on known topics that have been rarely researched to highlight topics with a high potential and identify institutions engaging in research activities related to those topics to show where potential innovators might be located. The feasibility to identify unknown topics from a TIM analysis only based on Scopus data is thus questionable. Many researchers interested in African agriculture tend to publish articles and research reports that are not referenced in journals. It is thus recommended to replicate the exercise by including recently published grey literature.

Number of publication records

One crucial aspect to be considered when utilising text mining methods is the statistical significance in the number of records being analysed. With the present example the number of records for the initial data set was sufficient enough for performing the data analysis and ensuring statistical soundness. However, if the topic turns out to be more niche and is highly specific, sufficient number of records may not be reached and obtaining keywords from the data set may not reliably reflect the key aspects of the topic.

Out in the field or staying in the office?

This exercise was presented at a workshop that was held at JRC premises on January 30th 2020 on "Horizon Scanning and Foresight Concepts, Tools and Showcases on Sustainable Use of Resources", where different concepts and horizon scanning approaches were demonstrated within both the European and international development context (see Workshop Agenda in Annex I & Workshop Briefing in Annex II). Two general methodological streams stood out from the discussions – desk-studies based on quantitative bibliometric data analysis (like in this exercise) and more qualitative approaches with expert gatherings in participatory workshop sessions. The latter one can encompass a fully enclosed horizon scanning process involving about ten to twelve experts in making sense sessions starting with exploring the topic, its facets, and possible development in the future, continuing with manually clustering the information and concluding on driving forces and paths for the way forward. On the other hand, as reported in this study, more automated desk-based approaches benefit from a more impartial and exhaustive list of academic records but cannot be stand alone and need to be complemented by expert judgments in various steps of the process. While the quantitative approach needs to have qualitative elements to make sense of intermediate outputs in the process, the qualitative approach could be strengthened by backing up, stimulating or initializing discussions with quantitative scientific evidence and facts from a quantitative analysis, meaning that a combination of both approaches could be the most efficient and persuasive way to implement a horizon scanning exercise.

We got the results – now what?

Recalling the purpose of horizon scanning as a first step to engage on strategic foresight and supporting anticipatory governance, it shall enable and convince stakeholders to consider weak signals, upcoming trends and challenges in their decision-making processes and to design solutions now to better prepare for the future. Therefore, a clear and concise communication and appropriate dissemination of the horizon scans is the next task that

should not be overlooked to reach out and convince decision makers (Cuhls, 2019). Outcomes can be synthesized as visual infographics, short on-line reports, or briefings to serve as input for political discussions. Ensuring an uptake of the generated information by the stakeholders largely determines the success and efficiency of a horizon scanning exercise. This proved to be often complicated due to the fact that identified signals, challenges and trends exceed the time horizon of short-term interests or may simply contradict with stakeholder's opinions, perceptions, interests, ideas, or plans.

5. Conclusion and Way forward

Combining Europe Media Monitor (EMM) and Tool for Innovation Monitoring (TIM)
As mentioned previously, adding grey literature and media coverage to the TIM based academic focus may be one solution to make the results more robust and updated in real time. One possibility could be to link TIM with JRC's Europe Media Monitor tool (EMM)[10]. EMM provides a set of tools to allow monitoring news items on a particular topic and customizing the search for specific organizations only, news websites, specific geographic locations and even including different languages. Though the data that is scraped from the web is usually not peer-reviewed and thus, not fully reliable and furthermore, not necessarily scientific, news data still offers a great potential for horizon scanning due to its timeliness and because it captures current interests and opinions on a specific topic. Similar to the previous exercise with academic records, a funnel approach could be implemented, gathering and clustering news items step by step from the more generic to the specific ones and using TIM to retrieve keywords reflecting a particular time period. Though processing retrieved articles for text mining purposes turned out to be quite complicated in the past due to aspects of copyright and intellectual property ownership, noteworthy, new regulations were adopted recently to cover legal aspects of copyright and data protection and thus allowing the analysis of news items by keeping them anonymous during the process. Last but not least, the expected noise in the collected articles could be filtered out by finetuning the news category and adding more keywords, but it most likely will remain a concern when applying text summarization methods like the TF -IDF algorithm.

Automatic clustering of information inside the data set
For the sake of optimization in terms of automation and timing, the workflow of the current horizon scanning process could be developed outside the TIM or EMM environments, through a programming language like python. Python offers numerous packages featuring endless opportunities for text mining and data analysis. After a data set is created inside python (Scopus library in python), a step like the clustering of keywords can be automated by clustering the publication records based on text similarity analysis of their abstracts. In a subsequent step, relevant keywords can be retrieved (e.g. by applying TF- IDF algorithm). Furthermore, as mentioned in chapter 2.3, abstractive text summarization methods based on machine learning or deep learning algorithms can be performed in order to get results on the content of the data set from a new perspective that cannot be achieved through usual extractive text summarization. While this approach ultimately offers the greatest flexibility for carrying out various insightful analytical and methodological approaches, it also requires the highest level of technical skills and knowledge.

Do we have the right data?
Conducting a horizon scanning in python, would be feasible with Scopus data that provide a useful API, but other data sources like PATSTAT or CORDIS would require a bit more effort and stronger IT related skills. Further, accounting for news items and grey literature may be

[10] EMM, see http://emm.newsbrief.eu/overview.html

an additional challenging task due to the lack of standard format, their noisiness and unstructured nature.
Nonetheless, the question of the right data is one that should be constantly and repeatedly raised as continuous advancements in the field of Natural Language Processing explore new data sources and make new data available, for instance in the field of speech recognition where data from speeches of policymakers or scientists from international conferences and other events can be translated to text and made available for further text mining.

Where are the weak signals?
The screening of keywords for potential signals is a challenging task and the most relevant keywords were not identified on the basis of the node size criteria. The list is ordered by node size and the node size is determined by TF – IDF scores (or simply by the term frequency as in the case of the top 300 authors). While some signals were identified somewhere in the top 10% of the list of retrieved keywords, others were also detected in the middle or even the bottom 10% of the list. This makes it impossible to set a specific numeric criteria or thresholds to shortlist keywords automatically as potential weak signals. Qualitative expert judgments remain a requisite for reviewing and screening the retrieved keywords.
A promising study conducted by JRC scientists succeeded with a quantitative semi-automatic approach to detect weak signals within the field of emerging technologies by scanning the "entire horizon" i.e. the whole list of publications in the Scopus data. Different indicators were developed and applied to each term of the corpus. An activeness score is assigning higher weights to weak signals that are featured in more recent publications, a patent indicator is checking how many patents compared to overall number of documents are contained in the signal and a hotness indicator that checks on the cross cutting nature of weak signals by looking at the number of different journals in which corresponding publications are featured and assigning higher weights to the signal if it is more debated in scientific conferences (Eulaerts et al. 2019). This approach would need to be tested when applied to a subset of publications targeting a particular topic instead of the entire corpus.

The unknown unknowns
When talking about horizon scanning and Foresight, practitioners always claim to strive for finding the unknown unknowns. What sounds like an exciting endeavour and a smart thing to do, turns out to be a rather utopian perception of what horizon scanning can deliver. Our knowledge provides the base for conducting these activities and thus it is certainly vain to search the unknown based on known patterns. However, constantly optimizing approaches in systematic iterations and building on both, qualitative and quantitative elements are certainly the best way to recognise these unknowns at the very moment when they step on the edge of our horizon and become something distantly known to us so our radars can detect them.
To conclude, our findings - including the outcome of the workshop (Annex II) - proved that horizon scanning should not be conducted as a standalone exercise, but rather should be integrated within a larger process of a foresight activity, where horizon scans are being interpreted and utilised respectively to transform potential signals into recommendations that can then be further articulated into tangible actions. This process cannot be carried out as a desk-based approach, but rather as a participatory process that accounts for an

appropriate engagement of relevant stakeholders already at an early stage of the process to foster the dialogue needed to reach tangible outcomes.

References

Abo-El-Wafa, H., Yeshitala, K. & Pauleit, S. (2017). Exploring the future of rural–urban connections in sub-Saharan Africa: modelling urban expansion and its impact on food production in the Addis Ababa region. In: Geografisk Tidsskrift-Danish Journal of Geography. Volume 117. Issue 2. July 2017. Pages 68 - 81.

Adenle, A.A., Wedig, K. & Azadi, H. (2019). Sustainable agriculture and food security in Africa: The role of innovative technologies and international organizations. In: Technology in Society. Volume 58. 2019. 101143. https://doi.org/10.1016/j.techsoc.2019.05.007.

Africa Agriculture Status Report (2018): Catalyzing Government Capacity to Drive Agricultural Transformation", Alliance for a Green Revolution in Africa, 2018, https://agra.org/wp-content/uploads/2018/10/AASR-2018.pdf.

Ahmed, N., Glaser, M. (2016). Coastal aquaculture, mangrove deforestation and blue carbon emissions: Is REDD+ a solution? In: Marine Policy. Volume 66. 2016. Pages 58-66.

Allahyari, M.; Pouriyeh, S.; Assefi, M.; Safaei, S.; Trippe, E. D.; Gutierrez, J. B. & Kochut, K. (2017a). Text Summarization Techniques. A Brief Survey.

Allahyari, M.; Pouriyeh, S.; Assefi, M.; Safaei, S.; Trippe, E. D.; Gutierrez, J. B. & Kochut, K. (2017b). A Brief Survey of Text Mining. Classification, Clustering and Extraction Techniques.

AUDA-NEPAD (2019). African Innovation Outlook 2019, AUDA-NEPAD, Johannesburg ISBN: 978-1-928527-21-3.

Bado, V.B. & Bationo, A. (2018). Chapter One - Integrated Management of Soil Fertility and Land Resources in Sub-Saharan Africa: Involving Local Communities. In: Advances in Agronomy. Volume 150. Pages 1 - 33.

Boelmann, E. & Telsnig, T. (2016). Technology Innovation Monitoring (TIM) for mapping emerging photovoltaics and offshore wind energy technologies. EUR 28259 EN, doi:10.2760/865392.

Callo-Concha, D., Jaenicke, H., Schmitt, C.B., Denich, M. 2020. Food and Non-Food Biomass Production, Processing and Use in sub-Saharan Africa: Towards a Regional Bioeconomy. Sustainability 2020, 12, 2013; doi:10.3390/su12052013

Cornock, O. and Mzali, S. (Eds, 2019). The Report Africa 2019. Oxford Business Group country reports. Bloomberg Terminal Research Homepage: OBGR‹GO›. https://oxfordbusinessgroup.com/sites/default/files/blog/specialreports/949525/africa_2019_special_report.pdf

Cuhls, K.E. (2019): Horizon Scanning in Foresight – Why Horizon Scanning is only a part of the game. In: FUTURES & FORESIGHT SCIENCE. DOI: 10.1002/ffo2.23.

Cuhls, K.; Giessen, A. & Toivanen, H. (2015). Models of Horizon Scanning - How to integrate Horizon Scanning into European Research and Innovation Policies. Technical Report. 56p.

Davison, J. (2018). Agriculture, Women and Land. The African Experience. Routledge. 286p.

Faloutsos, C. & Oard, D.W. (1995). A Survey of Information Retrieval and Filtering Methods. USA.

Fenollar F., Mediannikov O. 2018. Emerging infectious diseases in Africa in the 21st century. New Microbes New Infect. 2018 Sep 21;26:S10-S18. doi: 10.1016/j.nmni.2018.09.004. eCollection 2018 Nov.

Flaherty, K., Ayoola, G., Ogbodo, J. & Beintema, N. (2010). Nigeria. Recent Developments in Agricultural Research. Agricultural Research Council of Nigeria. Agricultural Science and Technology Indicators. Country Note. October 2010.

Förster, J.J., Downsborough, L & Chomba, M.J. (2017). When Policy Hits Practice: Structure, Agency, and Power in South African Water Governance. In: Society and Natural Resources. Volume 30. Issue 4. Pages 521 - 536.
Habegger, B. (2009). Strategic foresight in public policy: Reviewing the experiences of the UK, Singapore and the Netherlands, Futures vol. 42(1), pp.49-58.
Garnett, K. (2018). LIFE SMART Waste 2 The Horizon Scanning Toolkit. Tools for managing the future. Guidance for users. Cranfield University and Waverley Consultants. https://www.sepa.org.uk/media/367059/lsw-b4-horizon-scanning-toolkit-v10.pdf
Graff, G.D. and Pardey, P.G. 2019. Inventions and patenting in Africa: Empirical trends from 1970 to 2010. Volume 23, Issue1-2, March 2020. Pages 40-64. The Journal of World Intellectual Property published by John Wiley & Sons Ltd. https://doi.org/10.1111/jwip.12139
Hitchcock, R.K. (2019 Foragers and Food Production in Africa: A Cross-Cultural and Analytical Perspective. World J Agri & Soil Sci. 1(5): 2019. WJASS.MS.ID.000522.
Joanny, J.; Agocs, A.; Fragkiskos, S.; Kasfikis, N. & Goff, J. Le (2015). "Monitoring of Technological Development - Detection of Events in Technology Landscapes through Scientometric Network Analysis," Proc. ISSI 2015 Istanbul 15th Int. Soc. Sci. Inf. Conf., pp. 2–3.
Kaliisa, R., Picard, M. (2017). A Systematic Review on Mobile Learning in Higher Education: The African Perspective. In: Turkish Online Journal of Educational Technology - TOJET. Volume 16. Issue 1. January 2017. Pages 1 - 18.
Kinyua M; Diogo RVC; Sibomana J; Bolo PO; Gbedjissokpa G; Mukiri J; Mukalama J; Paul B; Sommer R; Kihara J. 2019. Green manure cover crops in Benin and Western Kenya - A review. CIAT Publication No. 481. International Center for Tropical Agriculture (CIAT). Nairobi, Kenya. 41 p. Available at: https://hdl.handle.net/10568/105923
Kyriakarakos, G., Balafoutis, A.T. & Bochtis, D. (2020). Proposing a Paradigm Shift in Rural Electrification Investments in Sub-Saharan Africa through Agriculture. In: Sustainability. Volume 12. Issue 3096. Pages 1 -19.
Le Picard, H. 2019. Gestion des dechets et production d'electricite en Afrique: l'incineration au service de la ville durable, Etudes de l'Ifri, ISBN: 979-10-373-0066-9. https://www.ifri.org/sites/default/files/atoms/files/lepicard_dechets_2019.pdf
Lin, J., Sun, X., Ma, S. & Su, Q. (2018). Global Encoding for Abstractive Summarization. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Short Papers), pages 163–169.
Makaya, E., Rohse, M., Day, R., Vogel, C., Mehta, L., McEwen, L., Rangecroft, S. & Van Loon, A.f. (2020). Water governance challenges in rural South Africa: exploring institutional coordination in drought management. In: Water Policy. Uncorrected Proof. Pages 1 - 22.
Manacorda, M. & Tesei, A. (2020). Liberation Technology: Mobile Phones and Political Mobilization in Africa. In: Econometrica. Volume 88. Issue 2. March 2020. Pages 533 - 567.
Manning, C. D.; Raghavan, P. & Schütze, H. (2008). Introduction to Information Retrieval. USA: Cambridge University Press.
Mars, M. (2013). Telemedicine and Advances in Urban and Rural Healthcare Delivery in Africa. In: Progress in Cardiovascular Diseases. Volume 56. Issue 3. November–December 2013. Pages 326-335.
Moner-Girona, M., Solano-Peralta, M., Lazopouloub, M., Ackom, E.K., Vallve, X., Szabó, S. 2018. Electrification of Sub-Saharan Africa through PV/hybrid mini-grids: Reducing the gap between current business models and on-site experience. Renewable and Sustainable Energy Reviews. Volume 91, August 2018, Pages 1148-1161. https://doi.org/10.1016/j.rser.2018.04.018

Moro, A.; Boelman, E.; Joanny, G. & Garcia, J. L. (2018). A bibliometric-based technique to identify emerging photovoltaic technologies in a comparative assessment with expert review. In: Renewable Energy 123, S. 407–416. DOI: 10.1016/j.renene.2018.02.016.
Nafi, E., Webber, H., Danso, I., Naab, J.B., Frei, M. & Gaiser, T. (2020). Interactive effects of conservation tillage, residue management, and nitrogen fertilizer application on soil properties under maize-cotton rotation system on highly weathered soils of West Africa. In: Soil and Tillage Research. Volume 196. Issue 104473.
OECD/FAO (2018), OECD-FAO Agricultural Outlook 2018-2027, OECD Publishing, Paris/Food and Agriculture Organization of the United Nations, Rome. https://doi.org/10.1787/agr_outlook-2018-en
Olanipekun, I.O., Olasehinde-Williams, G.O. & Alao, R.O. (2019). Agriculture and environmental degradation in Africa: The role of income. In: Science of The Total Environment. Volume 692. Issue 20. November 2019. Pages 60 - 67.
Qaiser, S. & Ali, R. (2018). Text Mining. Use of TF-IDF to Examine the Relevance of Words to Documents. In: International Journal of Computer Applications 181. DOI: 10.5120/ijca2018917395.
Shimeles, A., Verdier-Chouchane, A. & Boly, A. (2018). Building a resilient and sustainable agriculture in Sub-Saharan Africa. Palgrave Macmillan, Cham, http://dx.doi.org/10.1007/978-3-319-76222-7.
UNEP (2006). Africa environment outlook 2: our environment, our health. United Nations Environment Programme (2006). http://wedocs.unep.org/handle/20.500.11822/9626.
Sonnenschein,L., Etyang Tiberious, E. 2019 Dealing with the Sins of the Past and Work towards a Nature-Based Future: New Innovative Technologies for Sustainable Agriculture. World J Agri & Soil Sci. 2(1): 2019. WJASS.MS.ID.000526.
Swei, A., Couper, L.I., Coffey, L.L., Kapan, D. and Bennett, S. 2019. Patterns, Drivers, and Challenges of Vector-Borne Disease Emergence. VECTOR-BORNE AND ZOONOTIC DISEASES, DOI: 10.1089/vbz.2018.2432
Trolio, J., Eckman, M., & Mehta, K. (2016). Leveraging Globalization to Revive Traditional Foods. Journal of Sustainable Development, 9, 212.
UNEP (2013). Africa environment outlook 3: our environment, our health. United Nations Environment Programme (2013). .
Vasconcelos, M.J., Cabral, A.I.R., Melo, J.B., Pearson, T.R.H., Pereira, H.d.A., Cassmá, V. Yudelman, T. (2016). In: Mitigation and Adaptation Strategies for Global Change. Volume 20. 2015. Pages 1361 – 1383.

List of abbreviations and definitions

API	Application Programming Interface
CAADP	The Comprehensive Africa Agriculture Development Programme
CIAT	International Center for Tropical Agriculture
EMM	Europe Media Monitor
FAO	Food and Agriculture Organization of the United Nations
FP	Framework Programme
GDP	Gross Domestic Product
GMCC	Green Manure/Cover Crops
HIV	Human Immunodeficiency Virus
HS	Horizon Scanning
ICT4D	Information and Communication Technology for Development
IFPRI	International Food Policy Research Institute
JRC	Joint Research Centre
NEPAD	New Partnership for Africa's Development
NGO	Non-governmental Organization
NUTS	Nomenclature of Territorial Units for Statistics
OECD	Organization for Economic Co-operation and Development
R&D	Research and Development
SA	South Africa
TF - IDF	Term Frequency – Inversed Document Frequency
TIM	Tool for Innovation Monitoring
UNEP	United Nations Environment Programme

List of figures

Annexes

Annex I: Workshop Agenda

Horizon Scanning and Foresight
Concepts, Tools and Showcases on Sustainable Use of Resources

Workshop co-organised by EC JRC Dir. D and I
30th January 2020
(9:00-12:30-14:00-17:30)
JRC Ispra, Room 1302, Bld 101

Foreword and overall purpose:

This workshop aims at demonstrating various horizon scanning approaches and tools and forward looking exercises, which can be embedded in the JRC projects and activities for more anticipatory policy support.

The workshop focuses on the example of sustainable use of resources, embedding socio-economic dimensions, both in the European and international development context. The case studies will illustrate possible paths on how to approach horizon scanning and foresight in various geographic contexts – namely Europe, Africa and Asia.

This workshop is relevant to create a shared understanding of concepts (e.g., what are the differences between horizon scanning and foresight?) and know how (e.g., what are the tools available?). It also aims at facilitating discussions on the potential role and value added of foresight for current activities of the JRC in the context of the new EC agenda (Green Deal and Global) (e.g., how can the JRC generate new ideas coming from horizon scanning and then integrate them to develop forward looking, set strategies and build capacity?).

9:00 Welcome and ice breaker

Christine Estreguil (JRC/D6) and Maciej Krzysztofowicz (JRC/I2)
Workshop purposes + Tour de table

9:15 Concepts and tools for horizon scanning and foresight: what can we do and how?
Methods and tools for horizon scanning, JRC experiences (45 min incl Q&A)
Maciej Krzysztofowicz, Alberto Moro (JRC/Dir I.-I2 unit)
Aldo Podavini (JRC/Dir I.-I3 unit)
Christine Estreguil and Steve Borchardt (JRC/Dir D.-D6 unit)
Amalia Munoz Pineiro (JRC/Dir F. -F7 unit)
Horizon scanning toolkit and foresight concepts (30 min incl Q&A)
Kenisha Garnett (Cranfield University, School of Water, Energy and Environment)

10h45-11h30 Coffee Break including hands on tools session (Text Mining tools TIM, EMM)

11:30 Methods and showcases in European context

The EU Foresight system for the environment (FORENV)- methodology and case studies (30 min incl Q&A)

Manfred ROSENSTOCK (EC DG Environment, Environmental Knowledge, Eco-Innovation & SMEs (ENV.A.3)

Practical use of foresight for policy in Europe: implications of global megatrends at national and regional levels: the case study of Slovenia and Western Balkans (30 min incl Q&A)

Owen WHITE (Collingwood Environmental Planning, UK)

12:45-14:00 Lunch *including hands on tools session*

14:00 *Session continued*…Methods and showcases in European context

Horizon Scanning in Europe: the work of the EEA/Eionet foresight network (aka NRC FLIS - National Reference Centres on Forward-Looking Information and Services) (30 min incl Q&A)

Ana Jesus (European Environment Agency, NRC FLIS coordinator)

Sylvia Veenhoff (German Environment Agency, National scans on Sustainable Resource Use)

14:30 Methods and showcases in international development context

Foresight processes in Africa for the sustainable use of natural resources encompassing socio-economic dimensions (30 min incl Q&A)

Arthur Muliro (Deputy Managing Director, Society for International Development, Rome)

A foresight and innovation example at the nexus agriculture, environment and nutrition (30 min incl Q&A)

Oluwabunmi Alijore (International Centre for Tropical Agriculture, CIAT, a CGIAR research center)

A foresight process in Asia for shaping the future of food (30 min incl Q&A)

Kenisha Garnett (Cranfield University, School of Water, Energy and Environment)

16h15-16h30 Coffee Break

16:30 Parallel group sessions and conclusion on possible paths and ideas for the way forward

[Colleagues will be divided into 2 groups to discuss about whether, what methods, topics and possible paths to develop a HS/foresight research. One group will be focusing on Europe, one group on International Development. Both groups will be discussing the lessons learnt and way forward on methodological aspects]

Conclusions of each group will be presented to all audience on possible paths and ideas for the way forward.

17:30 Concluding remarks and follow-up

Christine Estreguil (JRC/D6) and Maciej Krzysztofowicz (JRC/I2)

18:00 Pick up Ispra JRC to hotel

19:30 Social Dinner at Restaurant Lido, Angera

Annex II: Workshop Briefing

Parallel session on "How to approach Foresight in International Development : lessons learnt and way forward", workshop 30th January 2020, 16:00-17:00

Participants: Christine Estreguil (chair, JRC D6 Unit), Steve Borchard (JRC D6), Claudia Capitani (JRC D6), Kenisha Garnett (Cranfield Univ, case study on the 'shaping the future of food' in Asia), Arthur Miluro (SID, case study on foresight processes in Africa), Oluwabunmi Alijore (CIAT/GCIAR, case study on future of Agricultural Research and Innovations in Africa)

Wrap up Notes

Four case studies were presented at the workshop: one on Text Mining with a focus on Agricultural R&D in Africa, one on food in Asia, one on Agricultural R&D in Africa, one on foresight processes in Africa. They all illustrated possible paths on how to approach horizon scanning and foresight in various geographic contexts in Africa and Asia. Speakers concluded on the specificity of foresight in international development concept, the importance of social process supported by science based facts, and that lessons could be learnt from experiences presented on the European Context.

Foresight in international (sustainable) development seems particularly associated with participatory approaches and scenario building. It cannot involved solely tools and outputs and should be focused on the **process** itself.

-Context is very important, which means on one hand having a good understanding of both the socio-economic and the cultural context. Local partners can help on this but it is important even external facilitator have a good knowledge of the intervention field. On the other hand, caring about the context means the capacity and flexibility of adapting the process itself to the local conditions (e.g. foreseeing potential language barrier, but also cultural or religious factors preventing equal accessibility and participation).

-The process is a chance for **learning**, social learning across multiple stakeholders from different background, education level, culture, age and gender. Also learning between the experts/facilitators and the participants in the process is an important outcome, for example through the exchange of data evidence and traditional knowledge.

- The process is based on **trust**, trust building is required prior and during the process itself. Moreover, the process can help in breaking barriers and building bridges between stakeholders from different fields or background who are usually in conflict. **By experience after conducting foresight exercises in different context in Africa, the future is a "space" where dialogue and conflicts resolution could be easier than in the "today" life.**

The case study from Thailand highlighted how the external facilitators could not influence too much the inclusion of different stakeholders because of rigid social rules at country level, and they had to step back (farmers and policymakers cannot sit at the same level, therefore some elements of the value chain were not represented in the process). On the contrary, a broader participation was possible in West and East Africa, maybe thanks to a better integration of the facilitators' team into the cultural context, e.g. due to long-term interactions and cultural affinity.

Participation always implies managing power dynamics, by putting in place adequate strategies or apply specific tools, to ensure a plurality of voices. Some people are reticent to speak out in public, or have difficulties in envisioning the future because focused on the present. For these and other challenges, having some champions in the group can be helpful.

Champions are usually “community” (in a broad sense) members who have learned by doing or have been trained.
Facilitation is key for engaging different kind of stakeholders, and it is important to encourage participation even of policy- and decision-makers, who ultimately are those who can actually change the rules of the game, and business sector, which is “casting the dice”. For example through desktop models or tools, the business sector voice could be overlooked. (How to ensure this is not happening, how to ensure data can be interpreted through the context reality, how to step from scanning to diving in?)

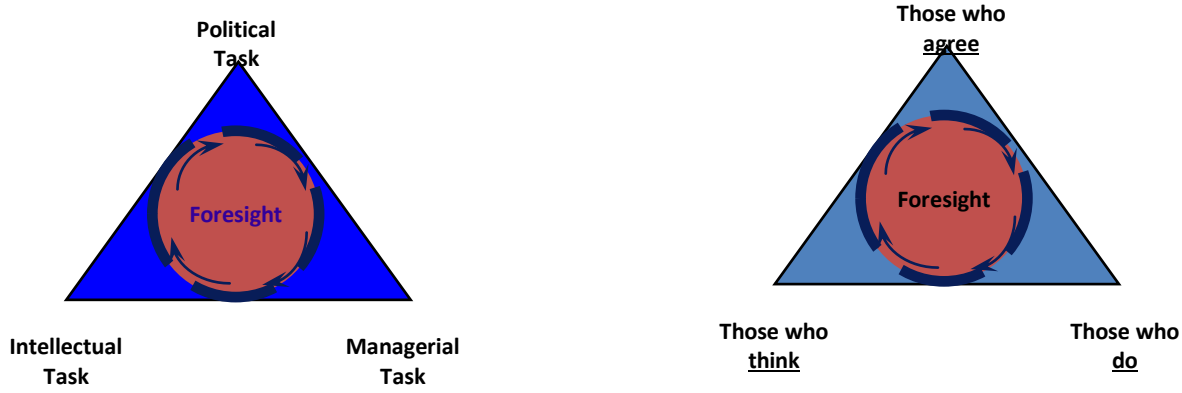
How can the JRC generate new ideas coming from horizon scanning and then integrate them to develop forward looking, set strategies and build capacity WITH stakeholders from emergent and developing countries in the context of sustainable development?
JRC >> AFRICA – **Facts check and data evidence from JRC can be an important support for the participatory scenarios**, e.g. to establish the context baseline, illustrate concrete and science based evidence, support the discussion process, exchange of ideas and reach consensus and/or better quantifying opportunities and challenges. Therefore, it is important to make JRC “data & tools” offer more visible and known and mainstream its utilization in public and private institutions. JRC expertise together with data and tools are perceived as a key input to facilitate and build discussions within Horizon Scanning/foresight participatory sessions. Those sessions would involve key stakeholders and JRC experts.
JRC>>AFRICA – Methodological steps could be as follow:
(1) Identifying **specific issues/topics in collaboration with JRC experts and stakeholders from emergent and developing countries** (for example ‘Active participation of African countries in R & D ?’ ‘More people empowerment in Development?’, ‘Any impulse of S&T, innovation systems in (agricultural) research ?’), or in alternative **start with a broad topic** (agricultural research and emerging signals),
(2) Running an **open and explorative process at JRC with Horizon Scanning tools** such as TIM and EMM. The aim would be to capture any signals, any ‘pockets of future’ in our present, from academics and media coverage, to be further proposed and discussed at participatory making sense sessions. These sessions could be also enriched by any available research facts and science based evidence to capture **‘the picture of now**’ and observed trends.
(3) Organise **making sense participatory sessions**(circa 12-15 participants per topic); they are means to channel and discuss JRC expertise, to **include stakeholders at DGs level, and have representatives from emerging and developing countries**. This participatory exercise is to be conducted preferably by an independent body outside the circle of policymakers. JRC could be a place to do so. The idea is to capture and set a common ground for dialogue on POSSIBLE/PROBABLE/REALISTIC/DESIRABLE/PREFERABLE ‘POCKETS OF FUTURE’ IN THE PRESENT, to discuss in an open manner possible stories for the future, priorities and (three) possible time horizons.
(4) **Build a foresight process to translate into actions, and influence our actions of ‘now’ for a better future.** JRC>>>AFRICA Horizon Scanning (HS) is a first step in the foresight process to help defining priorities, building strategies… HS gives a context to new signals, and is relevant as an input to anticipatory governance and periods of transitions. It brings more evidence, thus less debate... As a follow-up to HS, experts are then to be involved to **translate HS into actions, time horizons**, trends and scenarios, objective and thus define target.

(5) **Build a foresight process WITH stakeholders and the people in emergent and developing countries**

AFRICA>>JRC **Best strategic approach for JRC starting embarking on participatory approaches, either at region or country level**, depends on the existing relations which can ensure a quick win-win, and also on the topic. For example, existing regional networks or pilot countries where JRC has a role, could be used as entry points for wider regions or group of countries. **Some countries are considered as signposts to enter a region** (Kenya, Rwanda, Nigeria).

Particularly in international development context, Foresight is above all, a social process where the EC could play a role– it needs to get into the minds of decision makers: What are their assumptions? What are their values? What shapes and informs their views? How do you get them to appreciate things that are different from the 'official future'? There is a need to involve whole institutions and not just individuals in the process.

Foresight aims uniting all groups (triangle on the right) **in three policy tasks** (triangle on the left). Through foresight exercises, one can generate political will and dialogue, there is a need to share the learning and analysis, to address pardigms of social changes and define values to guide period of transitions.



GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:
- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
- at the following standard number: +32 22999696, or
- by electronic mail via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications from EU Bookshop at: https://publications.europa.eu/en/publications. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

KJ-NA-30349-EN-N

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub

Publications Office of the European Union

doi:10.2760/41501
ISBN 978-92-76-21446-5