agriculture

MDPI

Opinion

BeeOpen—An Open Data Sharing Ecosystem for Apiculture

Shreyas M. Guruprasad [1,†] and Benjamin Leiding [2,*,†]

[1] August-Wilhelm Scheer Institut für Digitale Produkte und Prozesse gGmbH, 66123 Saarbrücken, Germany; shreyas.mysoreguruprasad@aws-institut.de
[2] Institute for Software and Systems Engineering, Clausthal University of Technology, 38678 Clausthal-Zellerfeld, Germany
* Correspondence: benjamin.leiding@tu-clausthal.de
† These authors contributed equally to this work.

Abstract: The digital transformation of apiculture initially encompasses Internet of Things (IoT) systems, incorporating sensor technologies to capture and transmit bee-centric data. Subsequently, data analysis assumes a vital role by establishing correlations between the collected data and the biological conditions of beehives, often leveraging artificial intelligence (AI) approaches. The field of precision bee monitoring has witnessed a surge in the collection of large volumes of diverse data, ranging from the hive weight and temperature to health status, queen bee presence, pests, and overall hive activity. Further, these datasets' heterogeneous nature and lack of standardization present challenges in applying machine learning techniques directly to extract valuable insights. To address this issue, the envisioned ecosystem serves as an open and collaborative information platform, facilitating the exchange and utilization of bee monitoring datasets. The data storage architecture can process a large variety of data at high frequency, e.g., images, videos, audio, and time series data. The platform serves as a repository, providing crucial information about the condition of beehives, health assessments, pest attacks, swarming patterns, and other relevant data. Notably, this information portal is managed through a citizen scientist initiative. By consolidating data from various sources, including beekeepers, researchers, and monitoring systems, the platform offers a holistic view of the bee population's status in any given area.

Keywords: precision agriculture; data marketplace; apiculture; AI

check for updates

Citation: Guruprasad, S.M.; Leiding, B. BeeOpen—An Open Data Sharing Ecosystem for Apiculture. *Agriculture* **2024**, *14*, 470. https://doi.org/10.3390/agriculture14030470

Academic Editors: Francesco Marinello, Xiongzhe Han and Tianyi Wang

Received: 8 December 2023
Revised: 11 March 2024
Accepted: 13 March 2024
Published: 14 March 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Changes in temperature, humidity, illumination, and solar radiation can significantly affect honeybees, making them highly sensitive to fluctuations in environmental parameters [1,2]. Recent studies show that bees suffer from increasing stress factors, leading managed and unmanaged bee colonies to die or be less productive. In order to keep track of the health of their hives, beekeepers have traditionally depended on inspecting them manually and visually, a process that is both time-consuming and disruptive to colonies, resulting in infrequent assessments. Typically, beekeepers conduct hive inspections every two weeks during pollination or honey-production periods, with a reduced frequency during winter months [3]. However, with the advent of the Internet of Things (IoT) sensors and devices, passive monitoring of honey bee colonies has become feasible, giving rise to precision beekeeping applications [4]. These applications include monitoring hive health and colony strength through sensors measuring parameters like temperature, humidity, and CO_2 levels [5]. Additionally, audio and vibration monitoring has emerged as an effective alternative for remote hive surveillance, further expanding the options available to beekeepers [3].

The role of data in precision bee monitoring is pivotal for gaining insights into the behavior, health, and overall well-being of bee colonies [6,7]. Data collected from sensors

and observations provide valuable information about bee behavior, such as foraging patterns [8], hive activity levels [9], and swarming tendencies [10].

With the large-scale and heterogeneous nature of data, advanced data analysis techniques, such as machine learning algorithms, are required to recognize patterns, trends, and anomalies in bee behavior and health. Predictive modeling based on historical data can forecast potential hive issues, allowing beekeepers to implement preventive measures and optimize hive management strategies.

Both, the idea of digitizing apiculture [3,11], e.g., smart weight scales, acoustic sensors, temperature/humidity measurements and many more, as well as the application of AI in the field of apiculture [12,13] have been around for some time. Even industrial/commercial solutions have begun to emerge (https://amohive.com/, https://apic.ai/, https://bee-sensing.com/, accessed 8 December 2023). Similarly, apiary datasets and closed or (semi-)open data platforms for apiary data in general, as well as bee health in particular, exist. Even an EU-funded open data platform for bee health exists (https://bee-ppp.eu/, accessed 8 December 2023). However, the platform seems to be dysfunctional, and its dataset is very limited. Additional datasets exist, such as BeeInformed and others (https://beeinformed.org/, https://www.kaggle.com/datasets/se18m502/bee-hive-metrics, https://data.ontario.ca/dataset/honey-bee-pests-and-pathogens-in-ontario-apiaries, accessed 8 December 2023). However, none of them are updated regularly or allow collaboratively generating/curating/updating/sharing apiculture data. Moreover, the datasets are limited in size and variety, e.g., different sensors in combination with audio and or video material.

One possible solution to the challenge of sharing and reusing valuable yet scarce information is the use of data marketplaces [14,15]. However, this approach also has limitations, particularly when it comes to dealing with the issue of property claims erosion in a selling platform for data sets and models. Why do the stakeholders of the apiculture industry—and agriculture in a broader sense—hesitate to exchange data or work together to collect, share, collaboratively, and utilize data?

What is needed is an ecosystem that focuses on data acquisition, curation, and processing. This would enable the collaborative processing and sharing of diverse and broadly applicable AI models for apiculture, addressing the previously described needs. We advocate for a cooperative strategy that involves all stakeholders in the collection and analysis of apiculture-related data sets, along with the development of suitable AI models. This approach should be implemented within an ecosystem that offers incentives to all participants. The goal is to create a system that is transparent, secure, and incentivizes participation, while ensuring data integrity and accessibility. This manuscript describes the limitations of state-of-the-art collaborative data processing and sharing approaches to develop AI-driven applications for apiculture. Additionally, it explores the potential of utilizing a blockchain-based smart contracts.

2. Challenges of Data Sharing and Collaboration in Apiculture

The heterogeneous nature of these datasets and their lack of standardization present challenges in applying machine learning techniques directly to extract valuable insights. One of the primary obstacles in data collection within apiculture pertains to bias, as analyzing only a limited number of beehives over a fraction of bees' life cycles often yields insufficient insights [16–18]. To comprehensively understand bees, data collection must span an entire year, considering the significant variations in their behavior between seasons, especially between summer and winter, and even within the high season's onset and conclusion [19]. Consequently, amalgamating data from diverse apiaries and research initiatives becomes imperative to establish a robust dataset. This dataset enables model training on one subset and independent testing on another, mitigating the limitations of single-source data [18]. Ensuring the accuracy and reliability of data is essential for drawing valid conclusions. Inconsistent data collection procedures, sensor inaccuracies, and human errors can introduce noise and impact the quality of analyses.

The field of apiculture needs more standardized routines for data acquisition, curation, and processing, for the data itself and related metadata. In their work, Ugochukwu and Phillips [20] highlighted additional concerns that researchers have regarding data sharing in the field of plant phenotyping, which are also applicable to apiculture. These concerns include the potential risk of property claims being undermined, data misuse, and the absence of data harmonization.

An additional hurdle arises from the absence of publicly accessible datasets, codes, and trained models across many of the studies surveyed, impeding the ability to build upon prior findings. Consequently, there is a pressing demand for the establishment of a platform that facilitates standardized data collection and sharing within the research community. Such a platform would aggregate annotated data from diverse geographic locations, hive types, and honeybee sub-species, encompassing various recording conditions and quality standards. By advocating for harmonized data practices, this platform would drive advancements in precision beekeeping research and facilitate the benchmarking of results, thereby fostering collaboration and dissemination of knowledge within the field [17].

In [21], the authors list a collection of challenges that prevent participants of an ecosystem from openly sharing data. Thus far, the unresolved challenge lies in mitigating the risk of data loss through unauthorized duplication, misuse, or an inadequately incentivized system. Primarily, the willingness to exchange data is constrained by a lack of accountability regarding handling shared data by unfamiliar entities, such as competitors, as well as ambiguous legal circumstances concerning copyright-related matters. Ultimately, the exchange of digital information through collaboration often results in the formation of a participant network where a small number of individuals contribute the majority of the input. On the contrary, a significant majority of participants primarily derive advantages from the network without actively contributing to its maintenance and growth, exhibiting what can be referred to as a "free-rider" phenomenon [22].

Finally, little thought has been given to how to use the data in the context of AI-driven applications for apiculture, e.g., the relation between sharing data and contributing to the well-being of an ecosystem and benefiting from ML/AI applications utilizing the very same dataset as an incentive structure for collaborative open apiculture platforms.

3. An Open (Data) Marketplace Ecosystem

Recent years have seen an increase in the use of data marketplaces for sharing, exchanging, selling, and profiting from data, services, and goods, e.g., [14,15,23]—some of them operate on distributed-ledger technology and others without. The primary reason for this is primarily a result of their capacity to hinder unapproved entry to distributed data, responsibility within the framework of sharing and/or gaining access to data, and a range of motivation systems for data sharing with a focus on establishing shared and often decentralized records. Smart contracts can automate data transactions, access rights, and rewards distribution to data users and data collectors. By treating data as a tradable asset, each dataset can be tokenized, representing ownership or access rights. This also enables small-scale (micro-) transactions, allowing users to buy or sell portions of datasets.

Rather than a singular marketplace that could result in monopolies or oligopolies, such as the Android or iOS marketplaces [24], we suggest implementing a network of interoperable, open, and decentralized marketplaces, such as the InterPlanetary File System (IPFS) (https://ipfs.tech/, accessed on 8 December 2023), to store data off-chain and reducing storage overheads. The stored data are fragmented and encrypted across multiple nodes to improve security. Within this framework, stakeholders can select any marketplace while enjoying full access to the entire ecosystem. Stakeholders engage in spot trades to acquire or sell data and models, or opt for a subscription-based model to ensure continuous access. There are various market mechanisms that can be utilized to motivate individuals to participate in environmental preservation efforts through managing datasets or creating new AI/ML models.

Therefore, we suggest developing an open and collaborative framework to manage apiculture data sets, as demonstrated in Figure 1. The ecosystem's stakeholders consist of individuals and entities engaged in the utilization, provision, curation, and development of data, as well as creators of AI/ML models. A decentralized autonomous organization (DAO) governance model for the platform enables decision-making processes, allowing the stakeholders to also vote on critical issues. This helps create transparent policies for usage, contributions and rewards governed by community consensus. Data providers procure data and transmit them to the established data repositories. It should be noted that users have the capability to serve as data providers through the utilization of applications that utilize an artificial intelligence and ML model, simultaneously augmenting the existing data pool with newly added information. Data curators are responsible for the systematic handling of incoming data, including the elimination of substandard input and the incorporation of relevant metadata and labels as appropriate. Data curators conduct quality assurance assessments on data provided by fellow curators. Data curators validate the quality and authenticity of data.

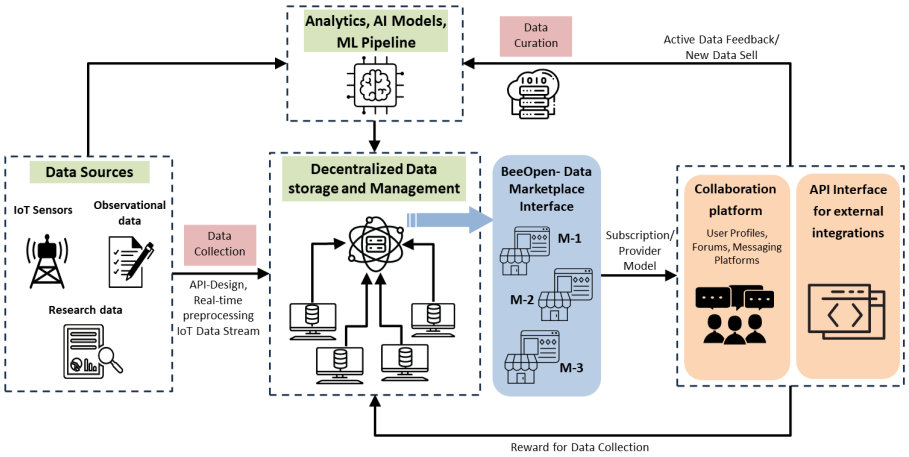

Figure 1. Main components of BeeOpen, enabling a collaborative and interactive open data sharing ecosystem, considering efficient data flow, user engagement, and incentive structure.

Further, the AI/ML trainers leverage ecosystem datasets to effectively facilitate the development, training, testing, and evaluation of AI/ML models, subsequently delivering them to the ecosystem. Customers have the option to perform single, non-recurring transactions (purchase/sale) of data or opt for a subscription-based model that provides ongoing access to a meticulously updated dataset and an artificial intelligence/machine learning model. Users can engage with a marketplace either through direct interaction or by utilizing an external user interface that leverages AI/ML model predictions derived from the underlying datasets.

4. Advantages for Modern Apiculture

In the context of apiculture, this pertains to the identification of a disease, such as American foulbrood, varroa mites, or colony collapse disorder, the prediction of honey production, or swarming activities, taking into account appropriately curated datasets based on the relevant geographical location. The data marketplace implements advanced search algorithms and filters to facilitate easy discovery of relevant datasets and a user-friendly interface for (non-)technical users. With regards to disease prediction, data providers indicate occurrences of diseases in certain areas, thus allowing for ML-based predictions on how they spread and advising nearby beekeepers on how to react appropriately. Other sensor inputs, e.g., audio or weight, to not only detect but also predict swarming behavior. Similarly, to the disease use-case, nearby beekeepers benefit from either nearby swarming

activities or swarming activities in regions/scenarios that closely resemble the conditions of individual beekeepers.

There are four distinct roles that have been identified, namely the collector, the ML modeler, the model user (such as the API of a mobile phone application), and the beekeeper. The data collector is responsible for compiling and organizing the datasets in accordance with established standards, metainformation, and data harmonization protocols before transmitting them to the designated data storage system. The model user compensates the data collector every time the ML model is utilized, either in whole or in part. The user model furnishes information to the beekeeper, for which compensation is rendered. Beekeepers contribute to data collection efforts by offering valuable reference data, such as disease information encompassing the position, incidence, and severity.

Subsequently, the impact on modern apiculture can be classified as the emergence of a potential role for beekeepers as providers of machine learning training data. By enhancing the availability and quality of training data, it is possible to enhance the effectiveness of forecast and prediction tools. Consequently, this improvement facilitates the enhancement of the dependability of apiculture forecast systems. As a result, beekeeping practitioners can effectively adjust their management routines and make informed decisions. The objective is to cultivate an engaged community within the ecosystem, supported by dynamic forums, assistance centers, and a wealth of educational materials.

5. Conclusions

We present an ecosystem-approach that facilitates scientists and industry stakeholders in the field of apiculture to enhance data accessibility. This approach introduces an open-data ecosystem that offers diverse incentives to its various stakeholders. It presents a potential approach for addressing the fragmented data storage and sharing across multiple islands, by establishing connectivity between modified datasets and metadata sourced from diverse stakeholders. It assists in addressing the existing limitation of machine learning models, whereby significant improvements in accuracy can be achieved by leveraging extensive sets of labeled training data. The roles and responsibilities of those involved in data collection, modeling, and the use of the model have been clearly outlined.

Author Contributions: All authors contributed equally to the writing of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: We acknowledge financial support by Open Access Publishing Fund of Clausthal University of Technology.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Abrol, D. Foraging in honeybees Apis cerana indica F. and A. dorsata F.(Hymenoptera: Apidae)-activity and weather conditions. *J. Indian Inst. Sci.* **1992**, *72*, 395.
2. Chen, W.S.; Wang, C.H.; Jiang, J.A.; Yang, E.C. Development of a Monitoring System for Honeybee Activities. In Proceedings of the 2015 9th International Conference on Sensing Technology (ICST), Auckland, New Zealand, 8–10 December 2015 ; pp. 745–750.
3. Abdollahi, M.; Giovenazzo, P.; Falk, T.H. Automated Beehive Acoustics Monitoring: A Comprehensive Review of the Literature and Recommendations for Future Work. *Appl. Sci.* **2022**, *12*, 3920. [CrossRef]
4. Meikle, W.; Holst, N. Application of Continuous Monitoring of Honeybee Colonies. *Apidologie* **2015**, *46*, 10–22. [CrossRef]
5. Cecchi, S.; Spinsante, S.; Terenzi, A.; Orcioni, S. A Smart Sensor-based Measurement System for Advanced Bee Hive Monitoring. *Sensors* **2020**, *20*, 2726. [CrossRef] [PubMed]
6. Zacepins, A.; Brusbardis, V.; Meitalovs, J.; Stalidzans, E. Challenges in the Development of Precision Beekeeping. *Biosyst. Eng.* **2015**, *130*, 60–71. [CrossRef]
7. Zacepins, A.; Stalidzans, E.; Meitalovs, J. Application of Information Technologies in Precision Apiculture. In Proceedings of the 13th International Conference on Precision Agriculture (ICPA 2012), St. Louis, MI, USA, 31 July–3 August 2012.

8. Baensch, S.; Tscharntke, T.; Ratnieks, F.L.; Haertel, S.; Westphal, C. Foraging of honey bees in agricultural landscapes with changing patterns of flower resources. *Agric. Ecosyst. Environ.* **2020**, *291*, 106792. [CrossRef]
9. Arribas, F.A.; Hortelano, M.R. An Internet of Living Things based device for a better understanding of the state of the honey bee population in the hive during the winter months. *Comput. Electron. Agric.* **2023**, *212*, 108026. [CrossRef]
10. Zgank, A. IoT-based bee swarm activity acoustic classification using deep neural networks. *Sensors* **2021**, *21*, 676. [CrossRef] [PubMed]
11. Bourouis, A.; Benahmed, T.; Mokeddem, K.A.M.; Benahmed, K.; Lairedj, A.S. IoT for Smart Apiculture: Issues and Solutions. In Proceedings of the 2022 3rd International Conference on Embedded & Distributed Systems (EDiS), Oran, Algeria, 2–3 November 2022; pp. 61–66.
12. Catania, P.; Vallone, M. Application of a Precision Apiculture System to Monitor Honey Daily Production. *Sensors* **2020**, *20*, 2012. [CrossRef] [PubMed]
13. Grammalidis, N.; Stergioulas, A.; Avramidis, A.; Karystinakis, K.; Partozis, A.; Topaloudis, A.; Kalantzi, G.; Tananaki, C.; Kanelis, D.; Liolios, V.; et al. A Smart Beekeeping Platform Based on Remote Sensing and Artificial Intelligence. In Proceedings of the Ninth International Conference on Remote Sensing and Geoinformation of the Environment, Ayia Napa, Cyprus, 3–5 April 2023; Volume 12786, pp. 92–99.
14. Ramachandran, G.S.; Radhakrishnan, R.; Krishnamachari, B. Towards a Decentralized Data Marketplace for Smart Cities. In Proceedings of the 2018 IEEE International Smart Cities Conference (ISC2), Kansas City, MO, USA, 16–19 September 2018; pp. 1–8.
15. Lawrenz, S.; Sharma, P.; Rausch, A. Blockchain Technology as an Approach for Data Marketplaces. In Proceedings of the 2019 International Conference on Blockchain Technology, Honolulu, HI, USA, 15–18 March 2019; pp. 55–59.
16. Sledevič, T. In Proceedings of the The Application of Convolutional Neural Network for Pollen Bearing Bee Classification, Vilnius, Lithuania, 8–10 November 2018 2019; pp. 55–59. [CrossRef]
17. Uthoff, C.; Homsi, M.N.; von Bergen, M. Acoustic and Vibration Monitoring of Honeybee Colonies for Beekeeping-relevant Aspects of Presence of Queen Bee and Swarming. *Comput. Electron. Agric.* **2023**, *205*, 107589. [CrossRef]
18. Hadjur, H.; Ammar, D.; Lefèvre, L. Toward an Intelligent and Efficient Beehive: A Survey of Precision Beekeeping Systems and Services. *Comput. Electron. Agric.* **2022**, *192*, 106604. [CrossRef]
19. Zacepins, A. Application of bee hive temperature measurements for recognition of bee colony state. In *Applied Information and Communication Technologies*; Latvia University of Agriculture: Jelgava, Latvia, 2012; pp. 16–221.
20. Ugochukwu, A.I.; Phillips, P.W.B. Data sharing in plant phenotyping research: Perceptions, practices, enablers, barriers and implications for science policy on data management. *Plant Phenome J.* **2022**, *5*, e20056. [CrossRef]
21. Paulus, S.; Leiding, B. Can Distributed Ledgers Help to Overcome the Need of Labeled Data for Agricultural Machine Learning Tasks? *Plant Phenomics* **2023**, *5*, 0070. [CrossRef] [PubMed]
22. Hales, D. From Selfish Nodes to Cooperative Networks-emergent Link-based Incentives in Peer-to-Peer Networks. In Proceedings of the Fourth International Conference on Peer-to-Peer Computing, Zürich, Switzerland, 27 August 2004; pp. 151–158.
23. Leiding, B.; Sharma, P.; Norta, A. The Machine-to-Everything (M2X) Economy: Business Enactments, Collaborations, and e-Governance. *Future Internet* **2021**, *13*, 319. [CrossRef]
24. Mason, P. *PostCapitalism: A Guide to Our Future*; Penguin: London, UK , 2015.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.