Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 3 (2024)

Animal Husbandry Risk Assessment Based on Data Analytics

Veska Gancheva[1], Stella Vetova[2], Kamelia Raynova[3]
[1, 2, 3] Technical University of Sofia, Bulgaria

Abstract:- This paper presents an approach and model for data integration and risk analysis in animal husbandry production. The proposed model for data integration and its components are described in detail, including the organization of the workflow, the modules, and their connections used for data exchange. Based on the presented workflow, experiments were conducted using statistical data on the population of domestic animals. The analysis of the results from these experiments reveals trends in the risk associated with animal husbandry risk analysis in livestock production and in particular disease risk assessment in cattle because of ingestion of chemical substances during forage cultivation. Based on year-by-year studies and quantitative metrics, it has been established that the model for processing and analyzing statistical data to assess the risk of cattle diseases from ingesting soil-borne chemical compounds is applicable in the field of animal husbandry. This model enables the tracking of trends in the use of artificial fertilizers and chemicals over a specified period, based on pre-collected statistical data.

***Keywords**:* Animal husbandry, data analysis, risk assessment, workflow.

1. Introduction

There is a growing need to automate data integration processes to go beyond simple data extraction and modification. Different databases store data in various formats, complicating integration due to high dimensionality and redundancies [1, 2]. Data integration involves filtering operations to remove duplicates, convert data, and manage it effectively. Common data integration models include Extract, Transform, and Load (ETL); Data Streaming Integration; Extract, Load, and Transform (ELT); Data Transformation; Data Virtualization; and Data Replication [3, 4].

In animal husbandry, risks are categorized as financial and business risks [5]. Business risks encompass production or profitability risks, marketing risks, institutional, political, and legal risks, human risks, and technological risks. Financial risks arise from borrowing to finance activities, leading to issues like variable loan interest rates, inadequate cash flow to repay debts, and changes in credit terms. Methods for analyzing and assessing these risks include machine learning to predict financial outcomes, statistical methods for financial risk analysis, and artificial neural networks for risk analysis and predictions.

To address risk management in husbandry, authors use portfolio theory to propose a mathematical model for optimal resource planning [6]. This model aims to minimize portfolio risk and maximize returns, representing risk as the standard deviation of returns. Experiments validate the efficiency of this approach. A study compares risk management strategies for insured and non-insured animals, highlighting effective strategies like vaccination, land ownership, access to financial resources, and livestock insurance [7]. Key risks include epidemic diseases, family health issues, climate and marketing risks, feed cost variability, and treatment delays.

For automating data collection and processing in animal husbandry, [8] propose a cloud livestock platform with a data centre, resource sharing platform, and business cloud service. This architecture accelerates the formation of digital farm profiles by collecting and integrating data directly from farms. They ensure data quality control using gross error estimation. A cloud-integrated farm management information system designed to connect and exchange data with other information systems is proposed [9]. This system offers data-collection and data-sharing services, demonstrating simplicity, flexibility, and reliability. Moreover, management technology in animal husbandry includes solutions for data privacy, sensor monitoring, remote control systems, offline maintenance, and cloud storage resources [10].

This paper aims to present an approach and model for data integration and a workflow for risk analysis in livestock production and in particular disease risk assessment in cattle as a result of ingestion of chemical substances during forage cultivation.

2. Materials and Methods

2.1 Animal Husbandry Data Analytics System

The proposed animal husbandry data analytics system architecture is presented in Fig. 1 and consists of three layers, each of which combines the tasks to be performed.

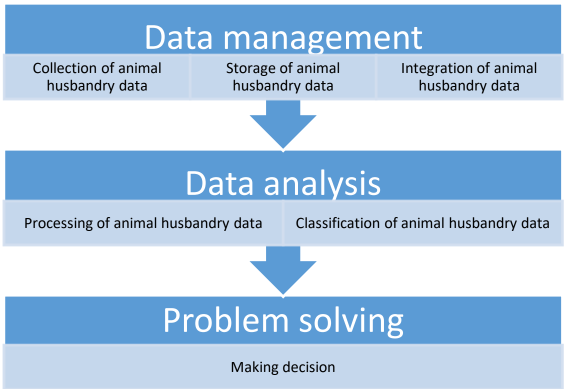

Fig. 1. System architecture for animal husbandry data processing

The proposed animal husbandry data analytics system architecture is presented in Fig. 1 and consists of three layers, each of which combines the tasks to be performed.

“Data Management” layer consists of three main phases: data preparation for analysis, interpretation and visualization, and the preparation phase includes animal husbandry data collection, animal husbandry data storage, and animal husbandry data integration. The "Animal Husbandry Data Collection" phase is based on the data sources.

The second phase "Animal husbandry data storage" of the proposed model is related to the data storage process. Typically, data is collected and stored in various file formats.

In third phase "Animal Husbandry Data Integration" process of merging data collected from different sources is carried out. That includes cleaning, mapping, and transformation steps. In the data integration process, data cleaning, unifying formats, grouping, filtering, downloading, loading from warehouses, extraction, merging, object data management. Storing the data in a data warehouse allows users to quickly access data from various sources, short time for information retrieval and storing large amount of data. This allows users to perform timely analysis and make predictions about future events and trends.

"Animal Husbandry Data Processing" phase of the second layer "Data Analysis" of the presented model includes the process and methods applied to process the animal husbandry data. The data processing process involves manipulating the collected data and performing functions and operations to extract meaningful information. Features included include validation, sorting, aggregation, analysis, reporting, classification. Sorting is used to arrange the data according to all submitted requirements. Aggregation is related to combining multiple datasets. Analysis is applied aimed to data transformation and modeling. Classification performs the separation of data into groups according to requirements.

The “Animal Husbandry Data Classification” phase involves arranging into groups data process based on predefined criteria. Clustering methods and techniques such as k-Means, k-Nearest Neighbor (kNN), Support

Vector machine (SVM), Artificial Neural Networks (ANN), Naive Bayes, Convolutional Neural Networks (CNN), etc. are applied for data purposes.

2.2 Dataset Selection

To carry out the experimental part of the study, a CSV file with statistical data on the amount used by artificial tools and chemical bond mentioned by the Food and Agriculture Organization of the United Nations is used. The file contains 141 records regarding the groups of chemical substances used for soil nourishment on the territory of the Republic of Bulgaria. The records used are represented in eight attributes as follows: domain code, domain, area code, area, element code, element, item code, item, year code, year, unit, value, flag, flag description. The data types are string and integer. The structure and contents of the CSV file are shown in Fig. 2. Record covers a period from 1990 until 2020. Statistical data provide information on the quantitative application of chemical fertilizers and compounds from the groups: Fungicides and Bactericides, Herbicides, Insecticides, Pesticides, Rodenticides and show their quantitative concentration in the soil as the main component in the soil-forage-cattle food chain in raising animals at home.

2.3 Animal Husbandry Predictive Risk Assessment Model

Fig. 3 presents the implementation of a model of the work process for the integration and analysis of data in animal husbandry and, in particular, the assessment of the risk of diseases in cattle as a result of the ingestion of chemical substances during the cultivation of feed. The constituents of the soil on which forage is grown for food fall into the soil-forage-cattle food chain. This has a beneficial effect on the development of various diseases in animal organisms.

1	Domain Cc	Domain	Area Code	Area	Element C	Element	Item Code	Item	Year Code	Year	Unit	Value	Flag	Flag Description
2	RP	Pesticides	100	Bulgaria	5157	Agricultural Use	1357	Pesticides (total)	1990	1990	tonnes	3906	E	Estimated value
3	RP	Pesticides	100	Bulgaria	5157	Agricultural Use	1357	Pesticides (total)	1991	1991	tonnes	3906	E	Estimated value
4	RP	Pesticides	100	Bulgaria	5157	Agricultural Use	1357	Pesticides (total)	1992	1992	tonnes	3906	E	Estimated value
5	RP	Pesticides	100	Bulgaria	5157	Agricultural Use	1357	Pesticides (total)	1993	1993	tonnes	3777.25	E	Estimated value
6	RP	Pesticides	100	Bulgaria	5157	Agricultural Use	1357	Pesticides (total)	1994	1994	tonnes	3648.5	E	Estimated value
7	RP	Pesticides	100	Bulgaria	5157	Agricultural Use	1357	Pesticides (total)	1995	1995	tonnes	3519.75	E	Estimated value
8	RP	Pesticides	100	Bulgaria	5157	Agricultural Use	1357	Pesticides (total)	1996	1996	tonnes	3391	E	Estimated value
9	RP	Pesticides	100	Bulgaria	5157	Agricultural Use	1357	Pesticides (total)	1997	1997	tonnes	3262.25	E	Estimated value
10	RP	Pesticides	100	Bulgaria	5157	Agricultural Use	1357	Pesticides (total)	1998	1998	tonnes	3133.5	E	Estimated value
11	RP	Pesticides	100	Bulgaria	5157	Agricultural Use	1357	Pesticides (total)	1999	1999	tonnes	3004.75	E	Estimated value
12	RP	Pesticides	100	Bulgaria	5157	Agricultural Use	1357	Pesticides (total)	2000	2000	tonnes	2876	E	Estimated value
13	RP	Pesticides	100	Bulgaria	5157	Agricultural Use	1357	Pesticides (total)	2001	2001	tonnes	2747.25	E	Estimated value
14	RP	Pesticides	100	Bulgaria	5157	Agricultural Use	1357	Pesticides (total)	2002	2002	tonnes	2618.5	E	Estimated value
15	RP	Pesticides	100	Bulgaria	5157	Agricultural Use	1357	Pesticides (total)	2003	2003	tonnes	2489.75	E	Estimated value
16	RP	Pesticides	100	Bulgaria	5157	Agricultural Use	1357	Pesticides (total)	2004	2004	tonnes	2361	E	Estimated value

Fig. 2. CSV file structure and content: attributes and first records representing breed trends for the study period

Talend Open Studio (TOS) was used to design the proposed workflow. TOS is an open source data integration software. Its functional architecture allows data analysis and review of analysis and query results when working with: databases based on Oracle, SQL Server, Postgre SQL; big data using Hive, Amazon-Redshift; delimited files. It enables data access and processing by connecting to the Apache Hadoop big data platform. TOS provides the user with the ability to process big data on a Hadoop cluster using big data jobs designed with Talend.

The implemented model is organized into four modules: SoilPolutiondata, FilterDataProcessing, SortProcessing, ReportProcessedData and the FilterProcessing, FDSortProcessing, ProcessedData connections used for data transmission. The model demonstrates the ability to load, process and analyze the 141 records initially defined in the original statistics CSV file.

The first phase of the integration process is to generate a metadata file for future use in the workflow, based on the original CSV file. The second phase is data filtering. It is implemented by the FilterDataProcessing module. Its purpose is to evaluate the raw data based on filtering criteria. The FilterDataProcessing module is the second in the workflow. Its structure is built using the tMap component. It is directly related to the input data where the original data was submitted. To analyze cattle disease risk, four fields were designed: "Element" with string data

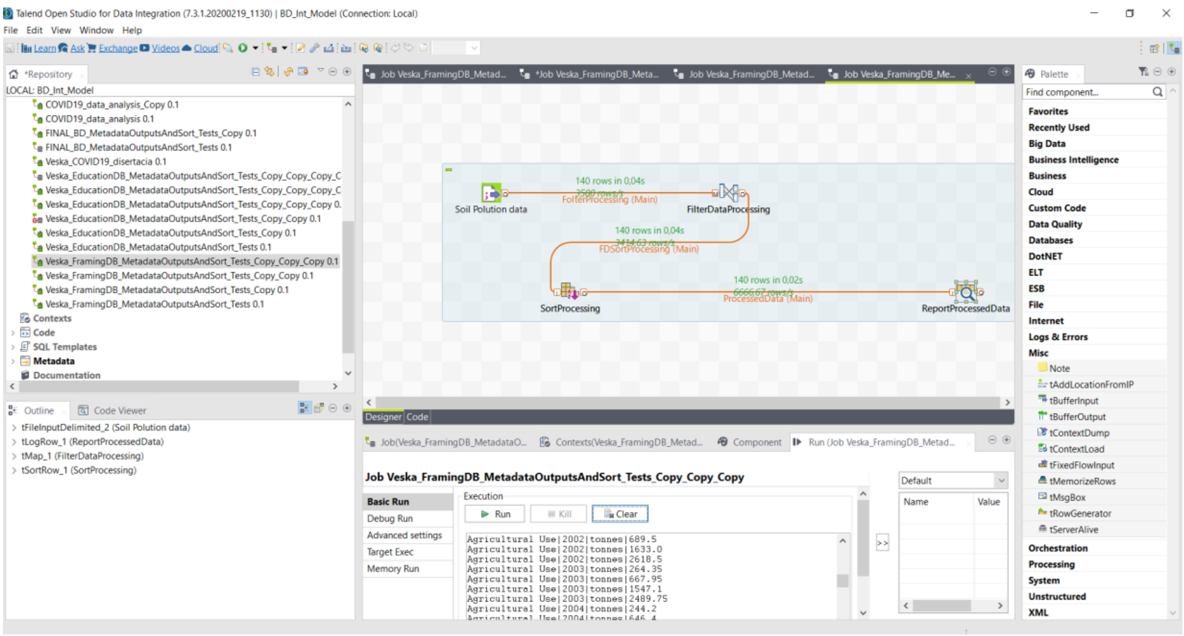

Fig. 3. Animal husbandry data integration workflow

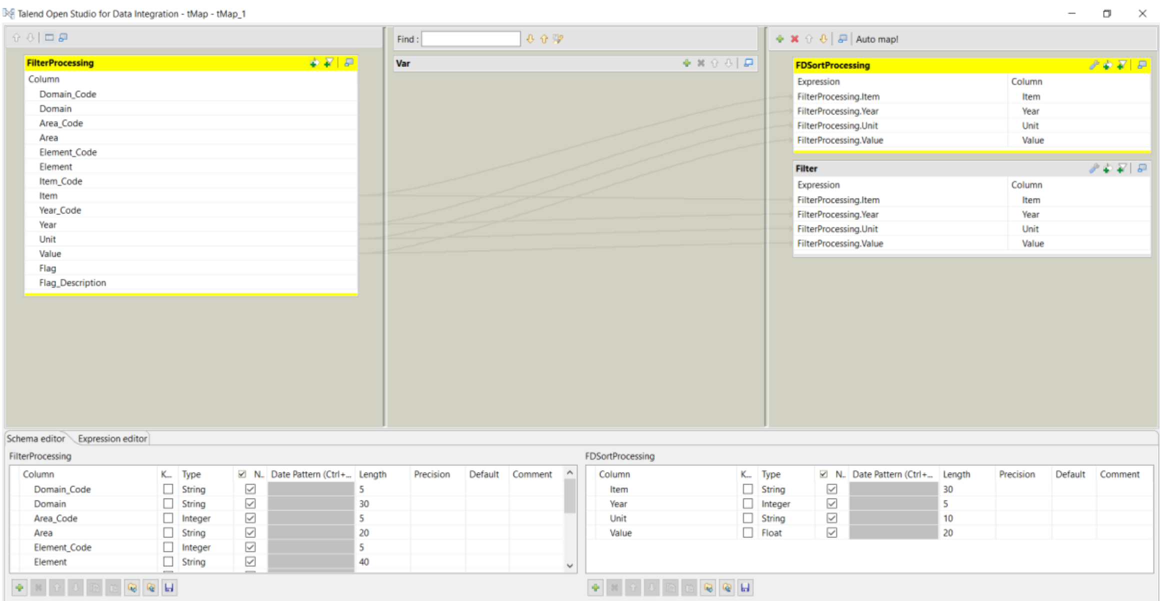

Fig. 4. Filter structure and projected fields

type; "Year" represented by Integer data type, "Unit" by String data type and "Value" by Integer data type. Fig. 4 shows the structure of the filter and the designed fields, including the data types used as described above.

The third phase is SortProcessing. It is designed using the tSortRow component to sort the result of the FilterDataProcessing filter by “Element”, “Year”, “Unit”, “Value” attributes. The sorting rules are structured in the following order:

- "Element" - ascending order;
- "Year" - ascending order;
- "Unit" - ascending order;
- “Value” - ascending order.

Fig. 5 shows the structure of the designed sorting rules by attribute, data type and sort order, and Fig. 6 illustrates the result of the whole workflow structure.

3. Results And Discussion

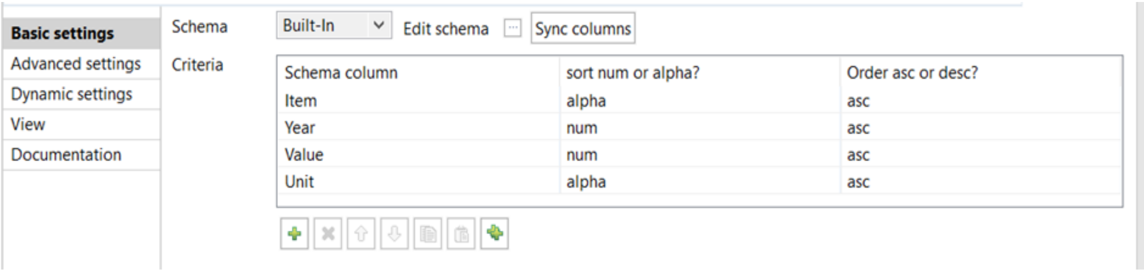

Fig. 5. Sort module model

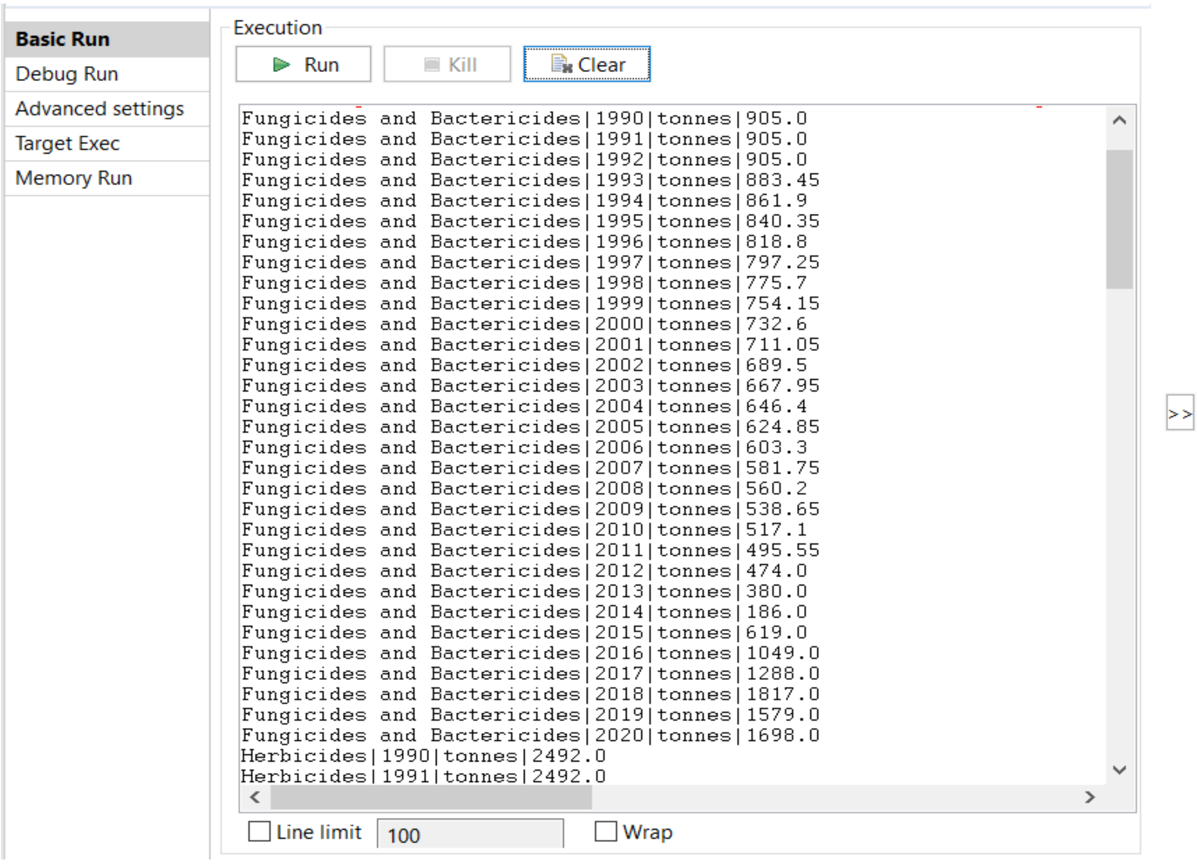

Fig. 6. Result of the workflow structure

According to the results of the data filtering and sorting operations, it is evident that data on the concentration of fungicides and bactericides were first reported in 1990, with a value of 905 tons. This trend remained stable until 1992, followed by a decline of approximately 100 tons for the periods 1993-1996, 1997-2001, 2002-2006, and 2007-2010. After these periods of gradual decline, the usage of fungicides and bactericides reached a minimum of 186 tons in 2014, followed by a rapid increase, peaking at 1817 tons in 2018.

A similar trend is observed in the quantitative growth of herbicide use from 1990 to 2020. Initially reported at 2492 tons in 1990 in the Republic of Bulgaria, these levels remained consistent from 1990 to 1992. From 1993 to 2015, there was a gradual decline of approximately 100 tons of chemical substances. By 2015, herbicide usage reached a minimum of 636.2 tons. In 2017, a significant increase of more than 300% was observed, doubling by

2019 to a maximum of 4340 tons, and then dropping again to 1988 tons in 2020, close to the 1699 tons recorded in 2017.

The trend in insecticide use shows a gradual decrease. Starting at 486 tons in 1990, there was a steady decline of about 20 tons per year until 2012, when usage reached a minimum of 83 tons. After this decline, there was a variable upward trend, peaking at 730 tons in 2019, a 900% increase from the 83-ton minimum in 2015.

Pesticide usage from 1990 to 2020 also shows variability. From 1990 to 1992, levels remained at 3906 tons. From 1993 to 2014, there was a steady annual decrease of 100 tons, reaching a minimum of 1001 tons in 2014. From 2015 to 2020, there was a significant increase, with a 50% rise in 2015 compared to the 2014 minimum, culminating in a nearly 700% increase in 2019, totaling 6663 tons.

Rodenticide usage is characterized by low application rates. The first quantitative data were reported in 1990 at 23 tons, remaining consistent for three years. Afterward, there was a gradual decline of about 2 tons, reaching 19.95 tons in 1999. Subsequently, the quantity reached a minimum of 0 tons from 2012 to 2019. In 2020, rodenticide usage increased again to 10 tons.

The analysis of the increase in the use of artificial fertilizers and chemical substances for the groups Fungicides and Bactericides, Herbicides, Insecticides, Pesticides, and Rodenticides shows a similar trend for the period from 1990 to 2020. For the three years from 1990 to 1992, the quantities of fertilization with bactericides, herbicides, pesticides, insecticides, and rodenticides remained unchanged. After this period, there was a controlled decline in the use of these substances, reaching minimum values between 2012 and 2014. From 2015 onwards, there were significant quantitative increases in the values of artificial fertilizers and ingredients used.

4. Conclusion

A workflow for integrating and analyzing data in animal husbandry, specifically for assessing the risk of diseases in cattle due to the ingestion of chemical substances used in feed cultivation has been designed and implemented. For the experimental part of the study, a CSV file containing statistical data on the quantitative use of artificial fertilizers and chemical compounds from 1990 to 2020, provided by the Food and Agriculture Organization of the United Nations, was utilized.

The implemented model is organized into four modules: SoilPollutionData, FilterDataProcessing, SortProcessing, and ReportProcessedData, with connections FilterProcessing, FDSortProcessing, and ProcessedData used for data transmission. The model demonstrates the ability to load, process, and analyze 141 records initially defined in the original CSV file. The first phase of the integration process involves generating a metadata file for future use in the workflow, based on the original CSV file. The second phase is the data filtering process, which follows defined rules allowing the tracking of trends in the quantitative changes in the use of various chemical compounds. Sorting, the second process in the workflow directly relates to the output data obtained after filtering and includes rules for data sorting.

Based on year-by-year studies and quantitative metrics, it has been established that the model for processing and analyzing statistical data to assess the risk of cattle diseases from ingesting soil-borne chemical compounds is applicable in the field of animal husbandry. This model enables the tracking of trends in the use of artificial fertilizers and chemicals over a specified period, based on pre-collected statistical data. Following the automated data processing and risk assessment, the next steps could involve measures to protect the soil, a crucial component in the soil-feed-cattle food chain.

5. Acknowledgment

The research leading to these results has received funding from the Ministry of education and science under the National science program INTELLIGENT ANIMAL HUSBANDRY, grant agreement n Д01-62/18.03.2021.

References

[1] Chen P., Zhang C, Data-intensive applications, challenges, techniques and technologies: A survey on Big Data, Journal of Information Sciences, 275:314–347, DOI: 10.1016/j.ins.2014.01.015.

[2] Paton N., etc. (ed.) Data Integration in the Life Sciences: 6th International Workshop, DILS 2009, Manchester, UK, July 20-22, 2009, Springer, ISBN-10: 3642028780, 2009.

[3] Julyeta P.A. RuntuweneIrene, Irene Tangkawarow, C T M Manoppo, Salaki Reynaldo Joshua, A Comparative Analysis of Extract, Transformation and Loading (ETL) Process, IOP Conference Series Materials Science and Engineering 306(1):012066, DOI: 10.1088/1757-899X/306/1/012066.

[4] Vetova, S., "Biomedical data integration and innovations concept", 47th International Conference Applications of Mathematics in Engineering and Economics 2021 (AMEE'21), 2021, Bulgaria.

[5] Aditto, S., Gan, C., Nartea, G. V. , 2012, 'Sources of Risk and Risk Management Strategies: The Case of Smallholder Farmers in a Developing Economy', in N. Banaitiene (ed.), Risk Management - Current Issues and Challenges, IntechOpen, London. 10.5772/50392.

[6] Gilbert, E., Meiklejohn, L., A comparative analysis of risk measures: A portfolio optimisation approach, Investment Analysts Journal, 48(3), 223-239, DOI: 10.1080/10293523.2019.1643128

[7] Khan, M., Chander, M., Bardhan, D., Risk perceptions and risk management strategies of insured and non-insured dairy farmers: A comparative study, Indian Journal of Animal Sciences 84 (6): 694–698, June 2014

[8] Wang, B., Qi, J., Research on the Intelligent Cloud Animal Husbandry Platform, Mathematical Problems in Engineering, vol. 2022, Article ID 4940511, 10 pages, 2022. https://doi.org/10.1155/2022/4940511.

[9] Kim, J., Lee, C., Baek, S., Rhee, J.,Open farm information system data exchange platform for interaction with agricultural information systems, Agricultural Engineering International: The CIGR e-journal 17(2):296-309, June 2015.

[10] Ivanov I, Vetova S, Stanchev G. Cryptographic Algorithm IDA for Security and Data Storage Increase in the Integrated Application of the European System "Ecall" For Automatic Emergency Calls in Car Accidents, IJEIT, Volume 4, Issue 12, June 2015, pp. 10 – 13.

[11] Talend Open Studio https://www.talend.com/products/talend-open-studio/