Animal Science Papers and Reports vol. 36 (2018), no. 2, 159-170
Institute of Genetics and Animal Breeding, Jastrzębiec, Poland

Design and implementation of a database enhancing the collection, management and analysis of data in an animal sciences project*

Joanna Pławińska-Czarnak[1], Luiza Ochnio[2], Joanna Zarzyńska[1], Janusz Bogdan[1], Jarosław Kaba[3], Alicja Majewska[4], Emilia Bagnicka[5]**

[1] Department of Food Hygiene and Health Protection, Faculty of Veterinary Medicine, Warsaw University of Life Sciences, Nowoursynowska 159, 02-776 Warsaw, Poland

[2] Department of Econometrics and Statistics, Faculty of Applied Informatics and Mathematics, Warsaw University of Life Sciences, Nowoursynowska 159, 02-776 Warsaw, Poland

[3] Institute of Epidemiology and Veterinary Management, Faculty of Veterinary Medicine, Warsaw University of Life Sciences, Nowoursynowska 159, 02-776 Warsaw, Poland

[4] Department of Physiology Sciences, Faculty of Veterinary Medicine, Warsaw University of Life Sciences, Nowoursynowska 159, 02-776 Warsaw, Poland

[5] Institute of Genetics and Animal Breeding Polish Academy of Sciences, Jastrzebiec, Postępu 36A, 05-552 Magdalenka, Poland

(Accepted March 6, 2018)

Laboratory and veterinary research often produces very specific data, which are difficult to collect and standardize, hence to interpret. The multi-stage character and complexity of our project, the variety of laboratory equipment used and the involvement of external service providers made it even more challenging to collect all data in a unified electronic form and store them in one place. A research project conducted on a living organism requires a holistic approach to the results of different health indicators, to the efficiency of each animal in a study group and to data analysis – so as to fully and correctly evaluate the state of health, or the stage of illness.

*This research work was supported by the grant 2011/03/B/NZ6/03711 from the National Science Centre (NCN) Poland and this work was financially supported by the Polish National Science Centre (grant 2013/09/B/NZ6/03514).
**Corresponding author: joanna_plawinska@sggw.pl

In order to facilitate the work of experts who evaluate the parameters and the research data obtained in a series of experiments, a database was designed and made accessible on-line to logged-in users. While designing this database, interesting solutions and ready-made functionalities were applied, so as to efficiently collect and display data. Database design tools such as MSSQL Express and MS Access were used. The MSSQL database was managed by the SQL Server Management Studio. To create and display data on the website, MS Excel, Visual Studio 2008 Express and IIS Express were used. This new data management system provides for an efficient storage of and insight into all data records, and can be taken advantage of in a variety of comprehensive scientific and research projects in the area of animal science especially veterinary researches.

KEY WORDS: array / database management/ goat / milk / RNA / RIN

Experiments on living organisms are often based on a number of mutually dependent measurements of various production, morphological, biochemical and health traits. The results of such studies are closely interlinked, and only if taken as a whole do they fully reflect the health and productivity of the animals or the influence a given factor has on them. It is difficult to capture all data within just one database, particularly if a variety of tools are used in the examinations and studies, so the results are often displayed in a multifaceted manner. Usually, the results are data printouts generated directly by the analyser, or additional researcher notes taken on an ongoing basis during the observation of a group of animals, and might include, e.g. body temperature, body mass, milk yield, atypical symptoms or behavioural changes indicating an imminent illness. To comprehensively evaluate the state of an animal's health, it is necessary to review an enormous amount of research data of various kinds across a period of time, so as to be able to distinguish a temporary weakness from chronic or growing lesions, or to identify a need to repeat the examination on the same samples or on new ones. Frequently, the readings are ambiguous, affected by the quality of the biological material, specimen contamination, or faulty device calibration. The main problem is to identify such shortcomings, which can be difficult when a researcher has to simultaneously deal with numerous animal experimental groups, their ongoing monitoring and the necessity to browse many tables containing analysis results. An additional issue is the necessity to consult vast numbers of data with fellow specialists working in remote scientific centres who need the data to be presented to them quickly and in an convenient form, especially when a prompt response is required.

Although there are many publications on the design and management of data storage and data sharing systems, as well as numerous repositories and specialised software to analyse and compare genome fragments of miscellaneous organisms [Chervenak *et al.* 2000, Barrett *et al.* 2012, Clark *et al.* 2013, Antell *et al.* 2014, Austin *et al.* 2015], no universal tool has so far been created that would sufficiently enhance the collection and storage of data. Researchers are left to their own devices, which naturally poses a risk of making mistakes, e.g. in the copying of results, particularly when the animal groups examined are numerous, and every animal has to be individually evaluated. All reports on bio-projects are focused on database

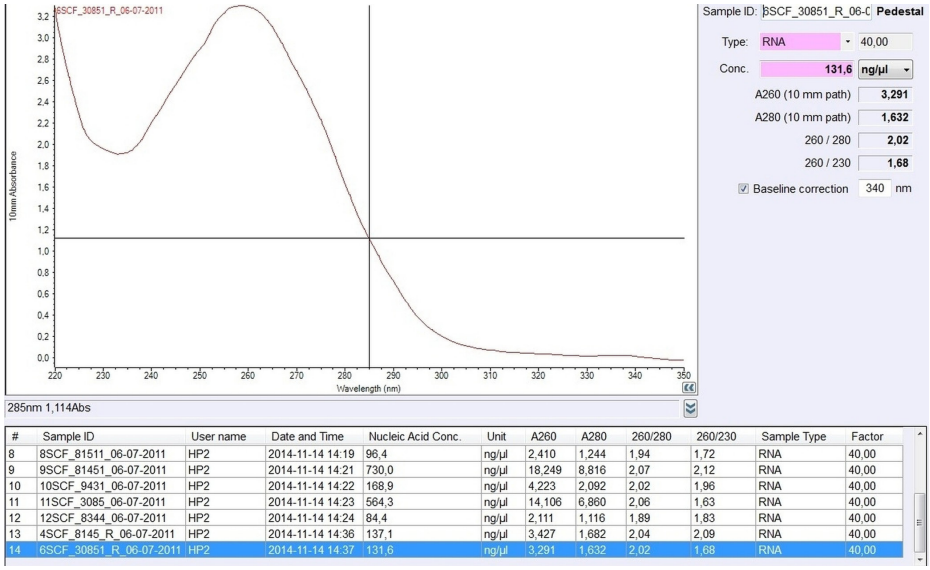
aspects with regard to the presentation and sharing of huge data volumes [Cole *et al.* 2014, Skoulis *et al.* 2015], on the automatic expert diagnosis of diseases of various animal species [Zeldis and Shawn 2000, Li *et al.* 2002; Zetian *et al.* 2005], or on the design descriptions of decision-making support systems [Kondratenko *et al.* 2015], rather than on highlighting the issue of structured data collection in various forms and formats. Therefore, of great value is every idea that would allow to enhance access to data and provide for better structuring and automation of data search and comparison of poorly collatable data.

This paper presents the design of another system to collate, store, search and process data with a functionality allowing various users to simultaneously view data sets. The system was adopted in a specific research project from the veterinary field, but the authors hope it might become an inspiration for scientists or veterinarians that are confronted with similar problems in their daily work.

Material and methods

The system has been design in purpose to be used as a tool for data collection, order, storage, archiving and analysis, to fulfil the needs of scientific projects based on animal models in the field of veterinary sciences. Goat milk and blood samples used for analyses were collected during research projects supported by two NCN grants. Between 2010 and 2015, blood and milk samples were collected five times per lactation cycle from goats of 2 breeds: Polish White Improved and Polish Colour Improved in the 1st, 2nd, 3rd and ≥4th lactation, from both healthy and small ruminant lentivirus-infected animals. The milk was physically and chemically tested to determine: fat, protein, lactose, casein, and urea content; dry mass; fat-free dry mass; freezing point; total acidity; density of free fatty acids; milk efficiency; and somatic cell count in 1 ml. Bacteriological examinations of the milk were performed in order to eliminate animals with udder inflammation. The project also encompassed cytometric examinations to determine the level of apoptosis in milk somatic cells, and the number of macrophages, lymphocytes and polymorphonuclear cells. RNA was isolated from the milk somatic cells, milk fat globules and goat peripheral blood nucleated cells. The isolated RNA was subject to further analysis for a number of necessary parameters regarding quantity, quality and integrity, which determined the suitability of the samples for use in transcriptomic examinations. Three times a year, blood samples were obtained from all the goats, and a blood serum analysis with the ELISA method was performed in order to detect antibodies to small ruminant lentiviruses that cause CAE (caprine arthritis encephalitis).

The analysis results came from various external laboratories and from a series of analyses performed on-site by a team of scientists and laboratory staff. The formula of the reading and storing of the data varied significantly, including hand-made notes, printouts with numbers and figures in all formats (e.g. PDF files), electronic recordings, or non-convertible analyser printouts (Fig. 1 and Fig. 2).



#	Sample ID	User name	Date and Time	Nucleic Acid Conc.	Unit	A260	A280	260/280	260/230	Sample Type	Factor
8	8SCF_81511_06-07-2011	HP2	2014-11-14 14:19	96,4	ng/µl	2,410	1,244	1,94	1,72	RNA	40,00
9	9SCF_81451_06-07-2011	HP2	2014-11-14 14:21	730,0	ng/µl	18,249	8,816	2,07	2,12	RNA	40,00
10	10SCF_9431_06-07-2011	HP2	2014-11-14 14:22	168,9	ng/µl	4,223	2,092	2,02	1,96	RNA	40,00
11	11SCF_3085_06-07-2011	HP2	2014-11-14 14:23	564,3	ng/µl	14,106	6,860	2,06	1,63	RNA	40,00
12	12SCF_8344_06-07-2011	HP2	2014-11-14 14:24	84,4	ng/µl	2,111	1,116	1,89	1,83	RNA	40,00
13	4SCF_8145_R_06-07-2011	HP2	2014-11-14 14:36	137,1	ng/µl	3,427	1,682	2,04	2,09	RNA	40,00
14	6SCF_30851_R_06-07-2011	HP2	2014-11-14 14:37	131,6	ng/µl	3,291	1,632	2,02	1,68	RNA	40,00

Fig. 1. Image from the initial quantity and quality analysis of RNA samples obtained from goat tissues. The screen display of an analysis performed by ThermoScientific NanoDrop ND-2000.

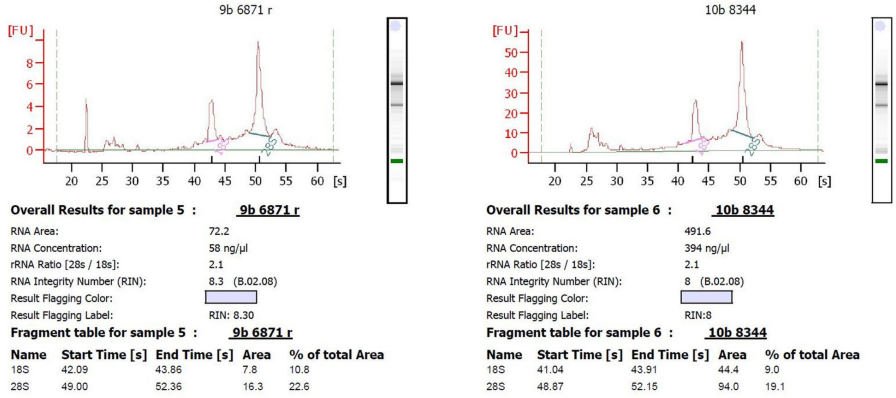


Fig. 2. Printout from the bioanalyser Agilent 2100; quantity parameters (RIN – RNA Integrity Number) of RNA isolated from two samples of goat peripheral blood nucleated cells.

Many analytical devices have their own software and data recording or report filing methods. Frequently, they use rare formats, an issue that prevents users from achieving adequate data consistency in one set.

In a research project conducted on a living organism, it is necessary to ensure a comprehensive overview of different health or efficiency indicators of particular animals in different test groups, and an analysis of the parameters obtained, so as to

duly assess the entire picture of the state of health or of the stage of illness. The ultimate goal is to achieve the complete comparability and integrability of miscellaneous data formats. The underlying idea was the maximum integration of the different data formats for the purpose of better data ordering and comprehensibility.

The authors undertook to design a database that would allow for an easier organisation and execution of such projects, owing to an enhanced access to complete information on every animal and its parameters, a functionality that would also promote more accurate sample selection for further analyses.

The following tools were used in the process of designing and implementing a system that would pool together all results as well as additional data for easier access to the information:

– data storage and access – MSSQL relational database;
– database management, structure building, viewing – SQL Server Management Studio;
– data import – Microsoft Access;
– designing software to extract analytical data from PDF files along with separate chart images – Visual Studio 2008 Express;
– in order to present the data and the results, a dedicated project website was created.

Table 1 shows the list of free software (with actual links) used for the database creation.

Table 1. The list of free software (with actual links) used for the database creation

Software and tools	Link
SQL Server 2017 Express is a powerful and reliable free data management system that delivers a rich and reliable data store for lightweight Web Sites and desktop applications.	https://www.microsoft.com/pl-pl/sql-server/sql-server-downloads
SQL Server Management Studio (SSMS) is an integrated environment for managing any SQL infrastructure, from SQL Server to SQL Database	https://docs.microsoft.com/en-us/sql/ssms/download-sql-server-management-studio-ssms
Visual Studio Community 2017 Free, fully-featured IDE for students, open-source and individual developers	https://www.visualstudio.com/downloads/
Internet Information Services (IIS) for Windows® is a flexible, secure and manageable Web server for hosting anything on the Web. From media streaming to web applications, IIS's scalable and open architecture is ready to handle the most demanding tasks.	https://www.iis.net/
ASP.NET is an open source web framework for building modern web apps and services with .NET. ASP.NET creates websites based on HTML5, CSS, and JavaScript	https://docs.microsoft.com/en-us/aspnet/#pivot=aspnet
PDFBox for .NET The Apache PDFBox® library is an open source Java tool for working with PDF documents.	http://www.squarepdf.net/pdfbox-in-net

Moreover, manual input data processing was performed and the results were presented using Microsoft Excel, whereas IIS Express was a tool used for the presentation of data on the website. The structure of the newly designed system builds on a database model executed in the form of tables storing formatted information. Since the main goal was to design a system ensuring user-friendly handling of information and easy implementation, a fairly complex source code was necessary. Another prerequisite was system management safety from the perspective of the system administrator that would eliminate the risk of unauthorised users introducing any changes. For this particular purpose, a table was created that stores information on the logging activity. Service access is based on authorisations in the database, a solution that promotes the management of authorisations to particular tables and views on the database level (Fig. 3).

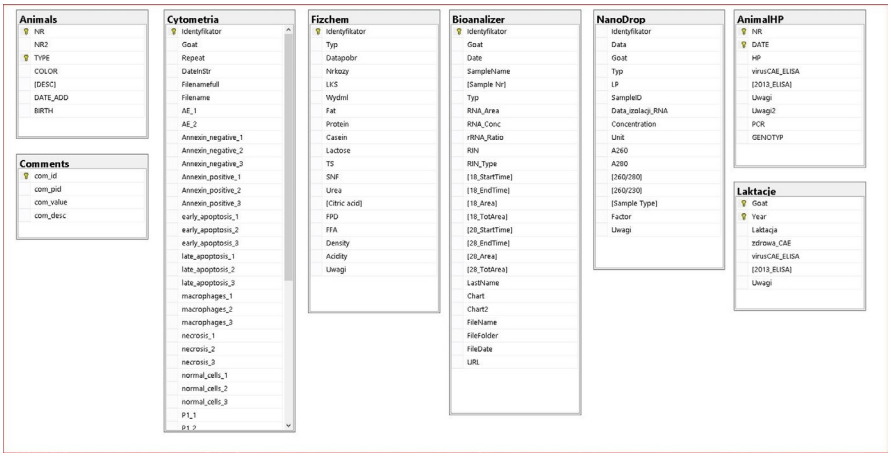


Fig. 3. Structure and tables of the relational database – image of tables containing research results, divided by source of information

Relational databases contained thematically ordered information regarding the animals and results of analyses.

Animals – the main reference table containing a list of all animals subject to the examinations; the identification number is the number assigned to a specific animal (unique record ID).

Tables containing data sets, depending on the type of analysis:

Cytometry – a table containing cytometry analysis results of particular cell fractions analysed in goat milk – individually for every animal and for every milk collection facility.

Fizchem – a table containing results of physical and chemical analyses of goat milk, somatic cell counts in 1 ml of milk (LKS) and milk efficiency.

Bioanalyzer – a table containing results of quality analysis of RNA obtained from isolated LKS and goat peripheral blood nucleated cells.

Nanodrop – a table containing data on the quantity of RNA in ng/µl and optical density OD, indicating potential organic or chemical contaminations in the RNA samples.

In every above-mentioned table, the identification key of the analysis is the animal ID number and the sample collection date.

Additional tables describing the examined animals:

Animal HP – defines the periods in which an animal is healthy or SRLV-infected.

Lactations – goat lactation number.

System functionalities and description

Similarly to many other systems based on World Wide Web Consortium standards (W3C – an organisation setting up standards on web page drafting and transfer) (Llorente, S. *et al.*, 2015), this system provides for access to database content using websites. It can be accessed through an internet browser. The system works with all popular internet browsers such as Firefox, Internet Explorer, and Opera, as well as with mobile devices using different operational systems.

Data can be downloaded by means of so-called database views, i.e. sources of data that are commonly used to create data sets easily presentable in MS Excel sheets, or that can be seamlessly transferred to webpages.

A collection of database views designed for the presentation of results is presented in Figure 4.

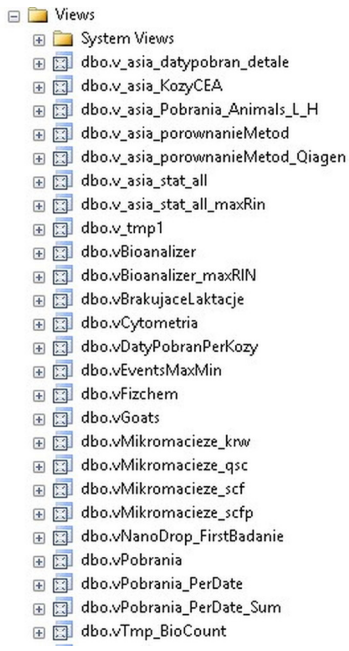
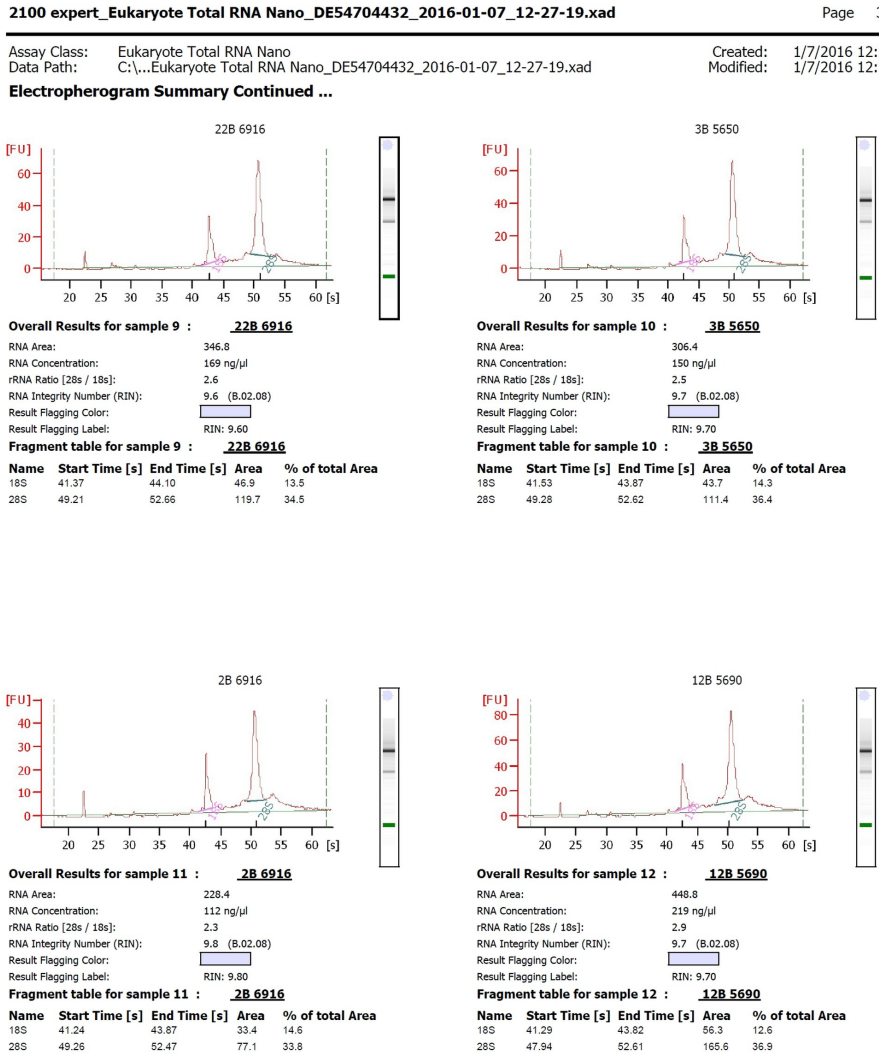


Fig. 4. List of so-called views in a database that enable further data processing and presentation, e.g. in an article or presentation.



2100 expert_Eukaryote Total RNA Nano_DE54704432_2016-01-07_12-27-19.xad Page 3 of 3

Assay Class: Eukaryote Total RNA Nano
Data Path: C:\...Eukaryote Total RNA Nano_DE54704432_2016-01-07_12-27-19.xad

Created: 1/7/2016 12:27:19 PM
Modified: 1/7/2016 12:50:26 PM

Electropherogram Summary Continued ...

Overall Results for sample 9 : 22B 6916

RNA Area:	346.8
RNA Concentration:	169 ng/µl
rRNA Ratio [28s / 18s]:	2.6
RNA Integrity Number (RIN):	9.6 (B.02.08)
Result Flagging Color:	
Result Flagging Label:	RIN: 9.60

Fragment table for sample 9 : 22B 6916

Name	Start Time [s]	End Time [s]	Area	% of total Area
18S	41.37	44.10	46.9	13.5
28S	49.21	52.66	119.7	34.5

Overall Results for sample 10 : 3B 5650

RNA Area:	306.4
RNA Concentration:	150 ng/µl
rRNA Ratio [28s / 18s]:	2.5
RNA Integrity Number (RIN):	9.7 (B.02.08)
Result Flagging Color:	
Result Flagging Label:	RIN: 9.70

Fragment table for sample 10 : 3B 5650

Name	Start Time [s]	End Time [s]	Area	% of total Area
18S	41.53	43.87	43.7	14.3
28S	49.28	52.62	111.4	36.4

Overall Results for sample 11 : 2B 6916

RNA Area:	228.4
RNA Concentration:	112 ng/µl
rRNA Ratio [28s / 18s]:	2.3
RNA Integrity Number (RIN):	9.8 (B.02.08)
Result Flagging Color:	
Result Flagging Label:	RIN: 9.80

Fragment table for sample 11 : 2B 6916

Name	Start Time [s]	End Time [s]	Area	% of total Area
18S	41.24	43.87	33.4	14.6
28S	49.26	52.47	77.1	33.8

Overall Results for sample 12 : 12B 5690

RNA Area:	448.8
RNA Concentration:	219 ng/µl
rRNA Ratio [28s / 18s]:	2.9
RNA Integrity Number (RIN):	9.7 (B.02.08)
Result Flagging Color:	
Result Flagging Label:	RIN: 9.70

Fragment table for sample 12 : 12B 5690

Name	Start Time [s]	End Time [s]	Area	% of total Area
18S	41.29	43.82	56.3	12.6
28S	47.94	52.61	165.6	36.9

2100 Expert (B.02.08.SI648) © Copyright 2003 - 2009 Agilent Technologies, Inc. Printed: 1/7/2016 2:36:40 PM

Fig. 5. Sample view of the bioanalyser PDF file. Every file contained an analysis of 12 samples, 4 samples per page, 3 pages in total. The figure presents the analysis results of RNA obtained from goat peripheral blood morphotic elements

Listings of particular groups of parameters in an MS Excel sheet are provided for easier execution of statistical analysis in respect of miscellaneous variability factors.

The usefulness of isolated RNA in array analysis or in Reverse Transcription Quantitative Real Time PCR (RT-qPCR) is dependent upon three basic parameters: quantity, which determines the number of tests planned, as well as purity and integrity, both of which determine the quality and credibility of genomic assays.

The quantity of nucleic acids is traditionally evaluated by UV absorption using a spectrophotometer. Currently, NanoDrop is a device which is widely used for prompt and accurate quantity measurements. A reliable method of measuring RNA quality is microcapillary electrophoresis and the analysis of electropherograms generated by, e.g. Bioanalyzer Agilent 2100. The bioanalyser software automatically generates the ratio of 18S and 28S ribosomal subunits, indicating the quality of the RNA (RIN = RNA Integrity Number). The bioanalyser generates results in the form of PDF files. The crucial challenge was to create a consistent mechanism to extract data from PDF files. The laboratory that performed those analyses could not submit the data in any other format. The files were collected over a long period of time, and their number at the start of the analysis ranged in hundreds. Every file contained 12 analysis results along with figures (Fig. 5).

For this purpose, a special program was created to analyse the indicated folder containing PDF files, with a view to create a CSV file with the results and separate charts as JPG files.

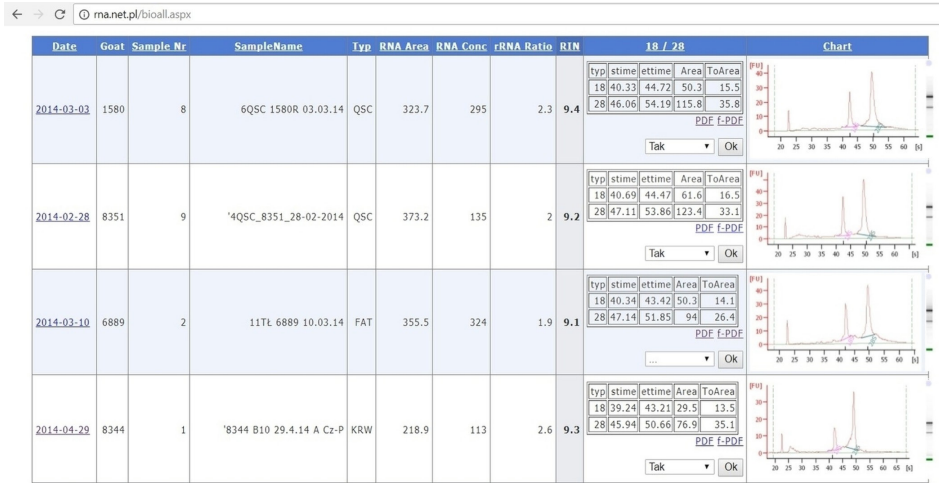


Fig. 6. Example: a screen displaying information, as presented on the website (https://rna.net.pl/; User: guest; Pass: rna1999), on individual animals (goats) from various sample collection times (Date), from three types of tissues (Type; QSC – milk somatic cells, FAT – milk fat, KRW – peripheral blood nucleated cells) and the marked RIN value for each sample (RIN) with the relevant chart (Chart). The image view can be modified by arranging the data on the website according to the following options: date, type of tissue, RIN value and the expert approval of the sample for subsequent analysis (Tak)

The program analyses dozens of files within seconds, building up a file ready for further manual processing that displays references to the JPG files with charts. The data obtained, after further processing (after the sample identification data in terms of tissue and collection time have been added), were loaded into the MSSQL database using Microsoft Access, software that offers user-friendly tools for data import (Fig.6).

Carefully selected unique keys assigned to the tables prevented the data from loading twice, e.g. due to the user's mistake, or due to the same file being sent twice to the data base.

Data storage, built on the free MSSQL database, made it possible to share the data on the internet, making it possible for many users to simultaneously download and edit the content. Every authorised user was given a login and a password, enabling them to view and structure data for further work.

Data displays arranged in this manner provided for a smoother sample selection at the most difficult stage of the analysis – building transcriptomic microarrays and verifying them by RT-qPCR. High quality samples were required (RIN $\geq$ 7.5), featuring sufficient concentration necessary for the planned number of assays. The selection of samples of the right quality guaranteed the correct execution of the laboratory testing and the credibility of the data obtained. Examples of the data sets used can be found in the open access repository Figshare (https://doi.org/10.6084/m9.figshare.5017148).

Conclusions

In biological or medical projects, it is of vital importance that the data are collected and stored in one place in a way making it possible for them to be read and modified by many users. Of great value in this respect are various databases that enhance the structuring, searching and processing of information. Nonetheless, there are no universal, ready-made software tools tailored for the needs of biomedical research. Each series of experiments has its own set of specific requirements and therefore calls for different solutions, particularly in the case of long term, multistage projects that last for years and encompass a wide spectrum of studies and analyses. The solutions outlined in this paper might present an example of how IT can support some types of experiments. It might be perceived as an idea, a model or an inspiration for researchers collecting and analysing data in similar projects. The documentation remarkably addressed the issue of data clarity; any abnormalities in examined animals were easier to detect, the health and condition profiles were easier to follow and interpret, the communication within the research team was leveraged, and it was possible to make references to tabularised content stored in one place. Structured and ordered information also brings other positive effects that became apparent during the project and that the authors of this paper have hereby attempted to highlight. The possibility to compare data that were initially completely incompatible allowed for a detailed statistical analysis, hence for revealing the presence or absence of various interrelations between them. For instance, it was observed that lactation,

milk efficiency or small ruminant lentivirus infection did not affect the quality of the isolated RNA (non-published data). The goat milk parameters were subject to analysis, and the results obtained served as a basis for a number of diploma theses in English [Gracey 2017, Jonzon 2017, Malmlöf 2017]. The clarity of comparison and easy access to data are factors that should encourage efforts to design the right place to store scientifically important information that is vital for researchers and tends to pile up as new projects are executed. The positive effects are: easier processing, enhanced data analysis and even a more enjoyable process of searching and supplying new information, owing to the visualisation of results. The system is an example of a "tailor-made" solution, adapted to the needs of a specific project in the field of veterinary science, but hopefully it can serve as an inspiration for future ideas to be developed by other researchers in response to similar problems faced during their scientific efforts.

REFERENCES

1. ANTELL K., FOOTE J.B., TURNER J., SHULTS B., 2014 – Dealing with data: Science librarians' participation in data management at Association of Research Libraries institutions**.** ***College & Research Libraries*** *75*(4), 557-574.
2. AUSTIN C.C., BROWN S., HUMPHREY C., LEAHEY A., WEBSTER P., FONG N., 2015 – Research Data Repositories: Review of current features, gap analysis, and recommendations for minimum requirements. ***International Association for Social Science Information Service and Technology (IASSIST)*** 39(4).
3. BARRETT T., CLARK K., GEVORGYAN R., GORELENKOV V., GRIBOV E., KARSCH-MIZRACHI I., KIMELMAN M., PRUITT K.D., RESENCHUK S., TATUSOVA T., YASCHENKO E., OSTELL J., 2012 – BioProject and BioSample databases at NCBI: facilitating capture and organization of metadata. ***Nucleic Acids Research*** 40, D57-63.
4. CHERVENAK A., FOSTERB I., KESSELMANA C., SALISBURY C., TUECKE S., 2000 – The data grid: Towards an architecture for the distributed management and analysis of large scientific datasets. ***Journal of Network and Computer Applications*** 23, 187-200.
5. CLARK K., PRUITT K., TATUSOVA T., MIZRACHI I., 2013 – BioProject. The NCBI Handbook [Internet]. 2nd edition. Bethesda (MD): ***National Center for Biotechnology Information*** (US).
6. COLE J.R., WANG Q., FISH J.A., CHAI B., MCGARRELL D.M., SUN Y., BROWN C.T., PORRAS-ALFARO A., KUSKE C.R., TIEDJE J.M., 2014 – Ribosomal Database Project: data and tools for high throughput rRNA analysis. ***Nucleic Acids Research*** 42, D633-642.
7. FIGSHARE https://doi.org/10.6084/m9.figshare.5017148
8. GRACEY Z.C., 2017 – Analysis of changes in the expression of selected genes regulating fat synthesis in goat's milk. Master thesis. Warsaw University of Life Sciences, Faculty of Veterinary Medicine, Department of Food Hygiene and Health Protection.
9. JONZON K., 2017 – Analysis of changes in the expression of selected genes regulating protein synthesis in goat's milk. Master thesis. Warsaw University of Life Sciences, Faculty of Veterinary Medicine, Department of Food Hygiene and Health Protection.
10. KONDRATENKO Y.P., KONDRATENKO G.V., SIDENKO I.V., KHARCHENKO V.S., 2015 – Model-oriented decision support system and fuzzy information processing for increasing efficiency of university-it-company consortia. ***Quantitative Methods in Economics*** 16, 157-166.
11. LI D., ZETIAN F, YANQING D., 2002 – Fish-Expert: a web-based expert system for fish disease diagnosis. ***Expert Systems with Applications*** 23, 311-320.

12. LLORENTE S., DELGADO J., MAROÑAS X., FLORIDO J., 2015 – Definition of standards-based building blocks for multimedia content management. ***Multimedia Tools and Applications*** 74(18), 7661-7683.
13. MALMLÖF M.N., 2017 – Lipid composition and fat globules relevance in goat milk during lactation cycle. Master Thesis. Warsaw University of Life Sciences, Faculty of Veterinary Medicine, Department of Food Hygiene and Health Protection.
14. SKOULIS I., VASSILIADIS P., ZARRAS A.V., 2015 – Growing up with stability: how open-source relational databases evolve. Information Systems, 53, 363-385.
15. ZELDIS D., SHAWN P., 2000 – Fish disease diagnosis program - problems and some solutions. ***Aquacultural Engineering*** 23, 3-11.
16. ZETIAN F., FENG X., YUN Z, XIAO-SHUN Z., 2005 – Pig-vet: a web-based expert system for pig disease diagnosis. ***Expert Systems with Applications*** 29, 93-103.