Business intelligence and data analytics framework: case study of humanitarian organisations refugees' registration system

Amal Ballout*, Mays Alhasan, Salam Fraihat and Ammar Elhassan

King Hussein School of Computing Sciences,
Princess Sumaya University for Technology,
Jordan
Email: amal.ballout7@gmail.com
Email: mays.alhasan@gmail.com
Email: s.fraihat@psut.edu.jo
Email: a.elhassan@psut.edu.jo
*Corresponding author

Abstract: Business intelligence and analytics has gained prominent focus among organisations with information systems that collect and process vast amounts of data. Voluminous, unprocessed data does not lend itself to offering useful insights for businesses, especially with basic statistical methods and traditional reporting techniques. In this work, we design a business intelligence and data analytics framework for refugee registration system serving over six million refugees to collect, collate and filter demographic data. The proposed reporting mechanism leverages the power of interactive dashboards to offer informative and intuitive reports and visualisations that are accessible and interpretable by stakeholders.

Keywords: business intelligence; analytics; demographics; reporting; dashboards.

Reference to this paper should be made as follows: Ballout, A., Alhasan, M., Fraihat, S. and Elhassan, A. (2022) 'Business intelligence and data analytics framework: case study of humanitarian organisations refugees' registration system', *Int. J. Innovative Computing and Applications*, Vol. 13, No. 1, pp.11–17.

Biographical notes: Amal Ballout is an MSc student specialising in data science at Princess Sumaya University for Technology, Jordan. She is also working as an IT specialist/consultant since 2016.

Mays Alhasan is an MSc student specialising in data science at Princess Sumaya University for Technology with industrial work in the same area. She specialises in: business and data analysis and visualisation, project management, process modelling and improvement, sales and market analysis.

Salam Fraihat is the Head of the Data Science Department at Princess Sumaya University for Technology. He has extensive teaching experience in computing, AI and data science, is very active in data science and BI research and coordinates cooperation and liaison between the university and the industry.

Ammar Elhassan is a Computing/Data Science Professor at Princess Sumaya University for Technology. He has worked in education management and quality as well as software development industry for over 20 years. Currently, he is active in developing BI and visualisation solutions to provide analytics to assess new learning approaches to cope with the impact of COVID-19.

1 Introduction

Technology analysts Gartner defined business intelligence (BI) as "an umbrella term that includes the applications, infrastructure and tools, and best practices that enable access to and analysis of information to improve and optimise decisions and performance" (Gartner, 2019).

Traditional reporting tools tend to be hindered by lack of flexibility and adaptability to fulfil the needs of decision makers in business and enterprise. BI framework consists of predefined report templates and analytics with filtering, drill-down, drill-across and slicing options that facilitate and support the mission of business users with minimal dependency on IT specialist skills and support. The domains

of application for BI tools are quite broad and include healthcare (El Morr and Ali-Hassan, 2019), education (Al Omari et al., 2019), finance (Sayedi et al., 2017), transport and risk management (Whiting, 2019) to name but a few.

Adoption rates of BI platforms within organisations have increased significantly in the past few years, this is partly due to BI technologies maturing and becoming more usable in organisations as a source of added value offering the organisation a competitive edge in access to insights that support decision making (Malladi, 2013). This, it is envisaged, will lead to improvements in performance and efficiency as well as data-driven strategic actions and business plans. Furthermore, non-profit organisations rely heavily on demographic historical and predictive statistics and analytics to secure funds for their programs and operations (IBM, 2019).

This article is organised as follows: Section 2 discusses the problem that triggered this project for humanitarian organisations registration. Section 3 shows briefly some related work. Section 4 emphasises on the value of the analytics and dashboards, Section 5 explains the methodology used in building this project; including the organisation operations and data understanding, requirements, architectures, data models, and dashboard, Section 6, shows the conclusion of this paper, finally a section that lists the references used in this paper.

2 Problem statement

Humanitarian agencies carry out direct relief and work programs for refugees from countries of war and crisis in difficult circumstances, terrain and against colossal challenges (Daar et al., 2018). Registered Refugees who receive support from these agencies need to be registered and databases need to be updated regularly in response to population changes such as those resulting from increases from births. Newly added descendants also need to receive care services including vaccinations (Lee et al., 2017) and food (Moffat et al., 2017).

It is, therefore essential to accurately and timely report on the number of registered refugees, based on different criteria, to help in planning and making decisions related to all programs that come within the organisation mission to help refugees and "consistent with internationally agreed goals and standards" (United Nations Relief and Works Agency et al., 2009). Technological developments play major roles in this area alongside BI, including Blockchain for connecting refugees and food programs (Zambrano et al., 2018), HCI technology (Talhouk et al., 2018) and also smartphones and devices (Dey et al., 2019).

3 Related work

Bulgarian Academy of Sciences applied BI technologies to get the most of their data and support decision making. Since the academy uses traditional classrooms for its education mission and utilises a mixture of documents in both paper and electronic format with some supplementary online information. Many reports and analytics are both challenging and time-consuming to generate, and most managerial decisions are made independent of the rich information pool available from students, classes, assessments, etc. Academy administrators saw an urgent need to transform data into information, knowledge and insights to support operations and decisions making processes. Therefore, a new, bespoke BI system was designed to cater for this very important requirement (Kabakchieva, 2015).

The Indonesian airline company saw the urgent need for a new solution to redeploy online transactional processing (OLTP) database which is used for daily operations including reservations and e-voucher payments, into a data warehouse that supports the OLAP analysis to retrieve informative analytics and reports in a user-friendly dashboard.

This project was implemented despite its complexity and huge deployment costs. Nonetheless, the added value contributed by the additional insights and information justified these overheads with significant improvements for the company in terms of better forecasting of demand, customer segmentation and determining the right price for each customer segment. Furthermore, this system helped in identifying potential target markets, and for product recommendations as well as marketing campaign planning and execution (Girsang et al., 2019).

4 Analytics and dashboards

Public sector and private sector organisations that offer social services often require deep analysis of demographic data for policy-making and strategic planning. Demographics included socioeconomic information about the population sliced by gender, age, ethnicity, housing characteristics, income, and so on. Demographic analysis provides an aggregate picture of the population, which helps in creating awareness about important population trends like birth rates, family size, ageing, poverty rate, etc.

These insights help in targeting social services such as pensions, education facilities, health services, housing planning and many other using demographic forecasts for the present and the future along with actual demographic data form a demographic information set (Sharan, 2018). Using this set along with related non-demographic information helps in making predictions about policy alternatives (Patil and Gangadhar, 2016).

Having deep understanding of the population demographics helps humanitarian organisations in planning resources efficiently. Some reports are already generated for this purpose, such as the Monthly Humanitarian Snapshot Report, Gender Bulletin, Regional Crises Emergency appeal, etc. (Girsang et al., 2019). These reports and other internal reports are important and provide good insights, but having all the needed charts and statistics in one place as a dashboard, will help widen the view and give more insights for different stakeholders and donors.

5 Methodology

5.1 Data understanding

Humanitarian organisations mandate is to provide different services to refugees and people of concern in its area of operations. These services include protection services, human development, humanitarian relief to name but a few. In order to pursue its mandate, those organisations have created specific criteria and standards to identify persons who are eligible to be part of its registration system to, subsequently, benefit from the agency's services. These standards and criteria aim to facilitate the agency's operations and services including health and education which are subject to budget limitations and relevant agency policies, regulations and bylaws.

The refugee registration system relies mainly on two entities: the individuals and their family. The Individual data includes all demographic information, and their registration information, while the family data links to the head of the family and his origin, the registration field, the current address and other details.

Registration data is stored in two data sources:

1 refugees registration SQL server database backups (2010–2018), these sources are available for BI solution

2 excel files of historical refugees statistics and reports before 2010.

The two sources are rich with reasonable quality for the BI application, but the only limitation is the different formatting for the historical files and the new schema found in 2010.

5.2 Requirements

5.2.1 High level business requirements

The BI solution should allow the head manager to control program operations resources, in terms of staffing, financing and services, to handle donors' advocacy and external reporting instantly with high accuracy and reliability, and to develop a centralised reporting and Business Intelligent tool for the Relief and Social Services Department (RSSD). In order to highlight most important figures, statistics and charts related to refugees registration in their fields of operations, also to provide a detailed Registration dashboard for top management which will support decision making, planning, monitoring, and management of different services the organisation provides such as health, education, protection and emergency cases.

5.2.2 Business processes

Propose a new reporting solution for the registration system with a more dynamic system supporting better aesthetics and user-friendly interface that helps control registration offices periodical restructuring to preserve the staff efficiency between 95%–100% with no lag in daily operations. In addition, to find the most vulnerable demographic groups based on their socioeconomic characteristics, services distribution planning, based on population density and other demographic characteristics. To continue the process of archiving the historical transactions of registered refugees based on their origin and other characteristics and to assess the rules of registration based on the density of population depending on their physical address vs. the registration address which is normally used, and finally to handle the emergencies and the related needs.

5.2.3 Business regulations and metrics

To monitor the registration processes in terms of: New dependents age groups, ratio of female to male refugees, areas and centres of population in relation to registration office capacities, and registered refugees vs. other-persons contrast, comparison and distribution. In addition, the substitution of the overwhelming number of individual ad-hoc requests from within the registration division, fields and other departments in addition to the external parties. For that purpose, data integration system needs to be efficient to extract the reports and charts in dynamic and timely manner.

5.2.4 Non-functional requirements

The system should ensure availability by providing standalone access to the BI tool. Portability through remote access from any device over the internet or mobile services is also a requirement. Data security by limiting access to privileged users only. Privacy by protecting and securing data according to predefined security settings. Flexibility through the ability to prompt ad-hoc inquiries on commonly used info, and the system should be well designed by having the appropriate visualisations and charts that provide highest effect of insights possible.

Additionally; easy accessibility and timeliness; by getting the desired results from the system in a timely manner and from the first attempt without an exhaustive workload, along with an easy to use and clear design that guides users to the required functionality without specialist support.

In order to be consistently up to date with latest bug fixes and enhancements, the system will be regularly and automatically uploaded with efficient uploading time of 30 seconds or less for all system components;/modules including the descriptive analytics and predictive analytics components.

5.2.5 Functional requirements

The main aspects of functionality offered by the system are:

1 Track refugee population growth overtime; compare population growth in five fields of operations; filter population decrement by event category (death, duplication or false registration); predict future population growth analyse post-emergency registration trends; and filter population increment by type of

registration (new dependents, reinstatements and new inscription).

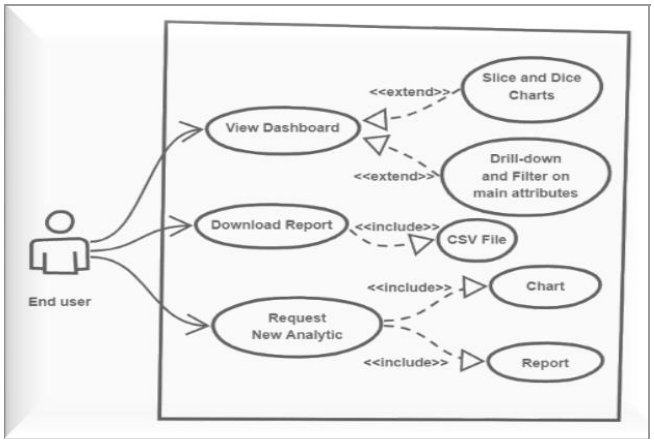
2 Extract refugee statistics; and categorise camp population based on demographic characteristics as well as by year of entry according to: age groups, registration status and registration field; and compare family registration addresses: registration address and physical address; additionally, calculate household size and composition, with the ability to filter by field, area or centre.

3 Analyse household heads in relation to: gender, marital status or age group. Analyse refugee registration status (12 categories) in terms of: number of registered persons per category, ratio of each category in each field and overall the five fields of operations; filter by nationality of registered person for each registration; extract number of refugees who were registered and never updated their registration profile in each field of operation.

4 Share BI documents and reports with co-workers inside the organisation via portal; view and download BI reports and share with external and internal stakeholders. Export BI reports data in commonly used standards (csv, pdf, xml, etc.); support drill-down, drill-across and slice-and–dice operations on different data components; and enable role-based access to the system by linking the access to the agency's active user directory.

Figure 1 represents a sample high level use case diagram, where the BI end user can login to the system, request a new analytic in a form of a report or chart, view a dashboard where they can slice, dice and drill-down and filter on mail attributes as well as download a report.

Figure 1 Use case diagram



5.2.6 Data requirements

In terms of data collection and engineering, the system requirements are as follows:

1 Consolidate data from disparate historical files and databases into one database.

2 Data integration should take into consideration the different types and formats of the two sources (see Figure 3).

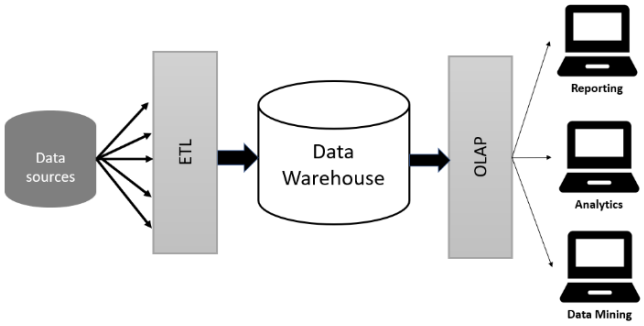
3 The data cleansing process needs to handle duplicates in refugee full name, mother's maiden name and date of birth. Random text and irrelevant values of the used attributes are to be replaced or cleared, depending on the type and the importance of these attributes. Moreover, null values in mandatory fields are to be replaced or excluded, depending on the attribute.

5.3 Architecture

The design of a BI system requires coherent strategic planning and effort. A framework that incorporates the best practices, standards and experiences is needed to start building the system and ensure that the desired BI system will be produced. BI systems consist of different components, but the key elements of any BI architecture are: data source, ETL process, data modelling, data warehouse, OLAP and tools and technologies (The University of New Hampshire Cooperative Extension, 2014). Figure 2 illustrates how these elements fit within BI system architecture.

The data source components cater for transactional processing as the input of the BI. This bulk of data feeds data warehouse with needed data, in addition to any other source of needed data such as benchmarking data and market research data. These various data sources could have different formats and standard, and should be addressed during the design process of the BI architecture (The University of New Hampshire Cooperative Extension, 2014).

Figure 2 BI architecture



The extract transform load (ETL) process extracts data from the operational data source and subsequently transforms and load sit into the data warehouse. This process can be completed using a script, or through various solutions on the market including Microsoft SQL service integration services and oracle data integrator. Data modelling helps specify data from the different data sources including data structures, formats, relationships and connections between data. Data warehouse represents a central repository of different data integrated from one or more contrasting sources. Stands for online analytical processing, it facilitates multi-dimensional data analysis (The Use of Demographic

Trends and Long-Term Population Projections in Public Policy Planning at EU, 2015).

Our proposed BI solution consists of two main components:

1 *Firstly,* the information and data architecture as shown in Figure 3. Data flows from data sources such as the Refugee Registration Information System (RRIS) and Historical documents, passing by data to a staging system where edits, changes and anomalies are managed, the data is then prepared, cleaned and transformed according to business requirements and finally be ready to be stocked into the data warehouse.

   From the data warehouse data can be integrated into an OLAP cube which stores and maintain data in multi-dimensional structure for analysis, querying, and reporting. The data then produce useful insights, metrics and KPIs in support of the Refugees Registration Division executives from different perspectives in a timely manner using BI application such as dashboard and reports.

Figure 3 Information and data architecture (see online version for colours)

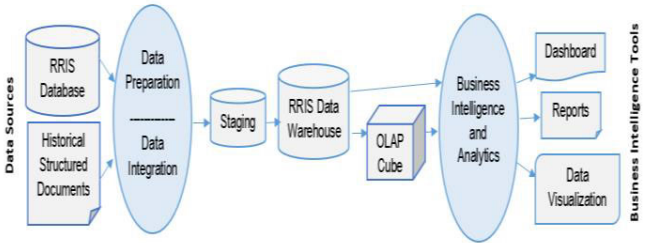
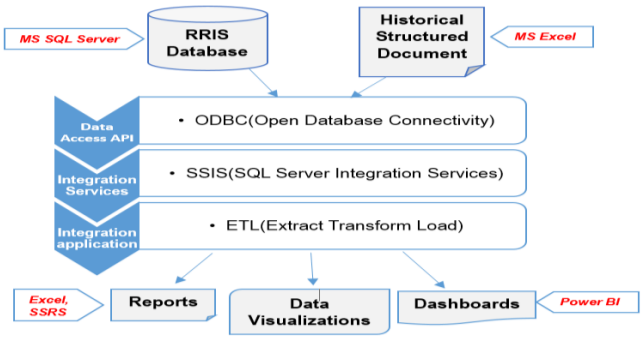


Figure 4 Technical and products architecture (see online version for colours)



2 *Secondly*, we have the technical and product architecture as shown in Figure 4. Here, the extract-transform-load process starts by extracting data from two distinct sources; i.e., SQL Server Databases and MS Excel files into the *staging* database in order to perform necessary processing and loading into the data warehouse.

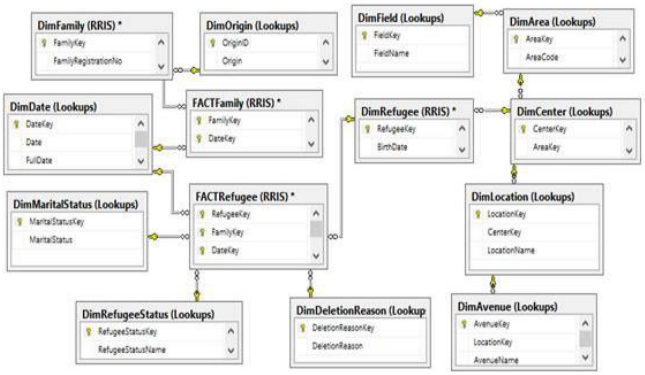
   The ETL integration service of choice is SQL Server Integration Services (SSIS) which is a platform that helps in data integration and data transformation processes. It utilises ODBC data access API's that work with SQL server databases and Excel file. The data is then made available within dashboard systems like MS power BI which enables reporting and exporting of insightful reports.

5.4 *Data models*

The proposed data model for the data warehouse shows the logical relationships and the data flow between different components of the RRIS system to be used in the BI solution.

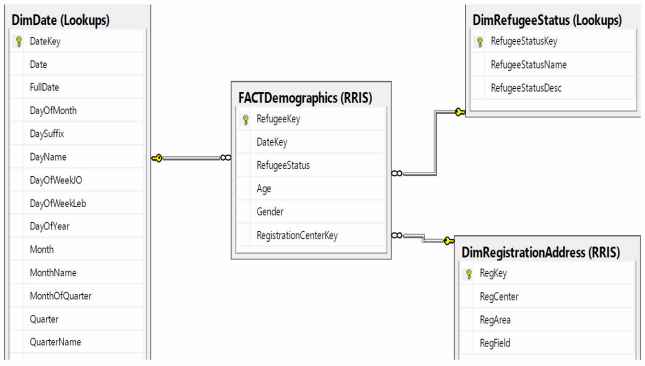
As shown in Figure 5, the main data model has Galaxy schema (Warnars and Randriatoamanana, 2016), which segregates data into two fact tables for the refugee and family data which is variable overtime. These two fact tables are associated with some common dimensions of the refugee information that are not subject to change.

Figure 5 BI data warehouse model (see online version for colours)



And the OLAP data model shown in Figure 6 is simpler and has big importance to form a multi-dimensional structure for reporting and analysis in timely manner. This model is used in the charts and visualisation that has high interaction with the user and may need to filter and drill down on different attributes, i.e., demographic attributes combined.

Figure 6 BI OLAP data model (see online version for colours)



5.5 *Dashboard*

Interactive BI Dashboards are gaining popularity in several industries and services with different size organisations deploying these solutions in support of executive departments.

To build a solid dashboard that covers all the system needs, MS Power Bi. Power Bi is a reliable, scalable, highly secure, cloud-based solution that is suitable for designers and data managers/analysts of different backgrounds and technical proficiency.

Power BI features are not limited to historical analysis, they also provide intuitive and powerful real-time and advanced analytics features as well. It also offers several distinct features such as automatic search for hidden insights, custom visualisations as well as straightforward integration with other Microsoft products in addition to the ability to connect with local/cloud data sources (Microsoft, 2019).

Because of the power of the visual elements, this visual analytics dashboard is very important and very informative for the organisation in general and for the registration division in particular. The dashboard and the related reports are designed based on the requirements and will be used as a platform to measure KPIs and metrics of the registration division, which will eventually affect decision-making.

As shown in F igure7, part 1 is a scorecard that shows the total population for the selected year, part 2, shows the Total population growth agency-wide with a prediction for the coming few years' totals, this can help in planning services and operations management on the ground. Part 3 shows the 'total population growth' for each field of operation separately, for comparison purposes with a caption that shows the main political events that mark the changes in the timeline. Part 4, is a slider that is linked to charts 2 and 3, and enables selection of the period of interest, which can be a single year or the full period from 1952 to the last year available in the data warehouse.

Figure 7 Refugees registration dashboard (see online version for colours)

Visualisations from 5 to 9 are all related to the demographics of the refugees, including gender, age, registration status, registration fields and areas. Part 5, is a pie chart that compares the population distribution based on gender for the selected year. Part 6, is a horizontal bar chart that compares the gender of the population by field.

Part 7 is a tree-map that illustrates the densest areas of registration across the five fields. Part 8 is a pie chart that compares the registration status of the whole population. Part 9 shows the distribution of the population by age.

Two origin maps are available, part 10 is a comparison between numbers of refugees based on their origins' districts between 1952 (start of operations), and their descendants' growth as of the selected year, plotted on a map. Part 11, shows the location of each origin (village or town) as of 1952.

Part 12 compares the number of refugees, based on their registration field (the place where the first ancestor refugee had registered), and the last updated address of those refugees and their descendants.

The current dashboard is just the beginning for a comprehensive, rich platform that will cover all business requirements for the refugee registration and eventually replace legacy reporting mechanisms to integrate reporting in a more convenient, visual and interactive way.

6 Conclusions

Many organisations around the world are facing challenges that could, potentially, be overcome with data analytics that allow these agencies to extract value from increasing data volumes. In the past decade many organisations adopted BI solutions to overcome traditional reporting system limitations and to provide better views of data including historical, current and predictive aspects of data related to the business operations. These views help in extracting insights and analytics in support of operations and decision making.

This paper has highlighted the potential value of BI and Analytics in relation to humanitarian organisations population demographics that are used for planning and decision making.

In the future, it is expected to expand the ideas presented here to gather data from other service/domain areas and create a comprehensive Business Intelligent system which can incorporate all refugee services with refugee demographics to better understand the trends and provide helpful insights to all stakeholders.

References

Ahn, N. et al. (2015) *The Use of Demographic Trends and Long-Term Population Projections in Public Policy Planning at EU*, October, National, Regional and Local Level [online] http://ec.europa.eu/employment_social/social_situation/docs/lot1_projections_summary_en.pdf (accessed 28 April 2019).

Al Omari, I., Al Omoush, R., Innab, H. and Elhassan, A. (2019) 'Visualizing program quality – a topological taxonomy of features', *2019 2nd International Conference on new Trends in Computing Sciences (ICTCS)*, Amman, Jordan, pp.1–10, doi: 10.1109/ICTCS.2019.8923062.

Daar, A.S., Chang, T., Salomon, A. and Singer, P.A. (2018) *Grand Challenges in Humanitarian Aid* [online] https://www.nature.com/articles/d41586-018-05642-8 (accessed 09 January 2020).

Dey, B.L., Babu, M.M. and Muhammad, S. (2019) 'Nature and outcome of technology adaptation: a study of Rohingya refugees' use of smartphones in Bangladesh', in *Academy of Marketing Conference*, July.

El Morr, C. and Ali-Hassan, H. (2019) 'Healthcare, data analytics, and business intelligence', in *Analytics in Healthcare*, pp.1–13, Springer, Cham.

French, C. (2014) *The University of New Hampshire Cooperative Extension*, November [online] https://extension.unh.edu/resources/files/Resource004765_Rep6784.pdf (accessed 27 April 2019).

Gartner (2019) *IT Glossary: Business Intelligence (BI)* [online] https://www.gartner.com/it-glossary/business-intelligence-bi/ (accessed 14 May 2019).

Girsang, A.S., Isa, S.M., Puspita, A.D., Putri, F.A. and Hutagaol, N. (2019) 'Business intelligence for evaluation e-voucher airline report', *Technology*, Vol. 10, No. 2, pp.213–220.

IBM (2019) *IBM Predictive Fundraising Analytics* [online] https://www.ibm.com/downloads/cas/Y8JKD0GO (accessed 15 May 2019).

Kabakchieva, D. (2015) 'Business intelligence systems for analyzing university students data', *Cybernetics and Information Technologies*, Vol. 15, No. 1, pp.104–115.

Lee, D., Weinberg, M. and Benoit, S. (2017) 'Evaluation of measles-mumps-rubella vaccination among newly arrived refugees', *American Journal of Public Health*, Vol. 107, No. 5, pp.684–686.

Malladi, S. (Ed.) (2013) 'Adoption of business intelligence & analytics in organizations – an empirical study of antecedents', *19th American Conference on Information Systems (AMCIS)*, Chicago, Illinois.

Microsoft (2019) *Why Power BI*, Microsoft Power BI [online] https://powerbi.microsoft.com/en-us/compare-power-bi-tableau-qlik/ (accessed 05 May 2019).

Moffat, T., Mohammed, C. and Newbold, K.B. (2017) 'Cultural dimensions of food insecurity among immigrants and refugees', *Human Organization*, Vol. 76, No. 1, p.15.

Patil, A.D. and Gangadhar, N.D. (2016) 'OLaaS: OLAP as a service', *IEEE International Conference on Cloud Computing in Emerging Markets*, pp.119–124.

Sayedi, M., Ghafari, P. and Hojati, E. (2017) 'Analysis of the effects and factors of implementing cloud business intelligence in banking systems', in *Proceedings of the10th International Conference on Utility and Cloud Computing*, December, pp.197–198, ACM.

Sharan (2018) *Business 'Intelligence Architecture – What, Why, and How'*, 30 November [online] https://analyticstraining.com/business-intelligence-architecture/ (accessed 29 April 2019).

Talhouk, R., Bustamante, A., Aal, K., Weibert, A., Charitonos, K. and Vlachokyriakos, V. (2018) 'HCI and refugees', *Interactions*, Vol. 25, No. 4, pp.46–51.

United Nations Relief and Works Agency (2009) *Consolidated Eligibility and Registration Instructions* January [online] https://www.unrwa.org/sites/default/files/2010011995652.pdf (accessed 27 April 2019).

Warnars, H.L.H.S. and Randriatoamanana, R. (2016) 'Datawarehouser: a data warehouse artist who have ability to understand data warehouse schema pictures', in *IEEE Region 10 Conference (TENCON)*, pp.2205–2208.

Whiting, B.D. (2019) 'Improved technical airworthiness taxonomy: capturing business intelligence to support an effective safety management system', in *AIAC18: 18th Australian International Aerospace Congress (2019): HUMS-11th Defence Science and Technology (DST) International Conference on Health and Usage Monitoring (HUMS 2019): ISSFD-27th International Symposium on Space Flight Dynamics (ISSFD)*, p.93, Engineers Australia, Royal Aeronautical Society.

Zambrano, R., Young, A. and Velhurst, S. (2018) *Connecting Refugees to Aid through Blockchain-Enabled ID Management: World Food Programme's Building Blocks*, GOVLAB [online] (accessed May 2019).