SUMMARY

The exciting field of the Internet of Things (IoT) is defined by Gokhale et al. (2018) as the internet connected with physical objects, and it enables various items such as electronic devices and sensors (wearable health monitors, connected appliances, smart home security systems, logistics tracking) and more to collect and exchange data.

The collected and exchanged data from IoT devices can be very large and heterogeneous; it is a substantial contributor to Big Data, which is why there is a strong tie between IoT and Big Data (Cecchinel et al., 2014, as cited in Zainab et al., 2015: 41).

Marjani et al. (2017) highlight the benefits for organisations, businesses and individuals originating from the ability to analyse and utilise IoT data; for example, smart cities, smart transport and smart factory equipment, among others.

However, raw IoT data is mostly useless without the ability to analyse it (Marjani et al., 2017). Therefore, Sarkar (2019) stated the importance of data processing; this set of tasks is known as Data Wrangling.

Huxley (2020) elaborates on the quality of quantitative data and the ability to make the data usable for planned analysis.

Messy data can be viewed as unorganised or unstructured data; because most analyses require a tabular format (Huxley, 2020). I agree with Sieminski (2023, as cited in Yair 2022) that messy data can mean many things more than the one above, and Yair (2022)

provides a more detailed definition. Sieminski (2023, as cited in Routhu 2022) states that sometimes messy data regarding structured data implies that the data needs to be transformed into an analysis-ready format; it may simply mean that the data must undergo a normalisation process.

Furthermore, data cleaning is crucial because otherwise, the risk is high that inaccurate results and conclusions are drawn from the analysis (Huxley, 2020).

Data Cleaning in the context of Big Data comes with limitations; it is time-consuming and requires domain expertise to achieve clean data (Ridzuan, 2019).

On the one side, humans outperform automated data-cleaning techniques (Geiger, 2023, as cited in Ding et al., 2022). On the other side, researchers, as mentioned in Geiger (2023, as cited in Shi et al. 2021), state that automated data cleaning tools improve data accuracy much faster than humans.

In Wong (2023, as cited in Horn, 2019) pointed out that humans are more likely to make errors than in automated data-cleaning processes.

Khalid (2023, as cited in Bhatt, 2020) pointed out that this time constraint mentioned above can be overcome by using data cleaning services, for example, by companies such as Experian (2021).

Tankard (2012) highlights the security issues in the context of Big Data since the data is primarily centrally stored in one place for analysis; data could be disclosed by human error

or through attackers. Therefore, the data needs to be stored safely and protected, and all regulations need to be met (Tankard, 2012).

REFERENCES:

Bhatt, V. (2020) The significance of data cleansing in Big Data explained, AiThority. Available from: https://aithority.com/guest-authors/the-significance-of-data-cleansing-in-big-data/ [Accessed: 11 February 2023].

Cecchinel, C. Jimenez, M. Mosser, S. & Riveill, M. (2014) An Architecture to Support the Collection of Big Data in the Internet of Things. IEEE World Congress on service: 442-449. DOI: 10.1109/SERVICES.2014.83

Ding, X. et al. (2022) IoT Data Cleaning Techniques. A Survey. Intelligent and Converged Networks 3(4): 325-339. DOI: https://doi.org/10.23919/ICN.2022.0026

Experian (2021) Case study: Royal Trinity Hospice - saves time and money with data cleansing, Experian UK. Available at: https://www.experian.co.uk/blogs/latest-thinking/marketing/case-study-royal-trinity-hospice/ [Accessed: 11 February 2023].

Geiger, J. (2022) Collaborative Learning Discussion 1: Peer Response Available from: https://www.my-course.co.uk/mod/forum/discuss.php?d=135855 [Accessed 11 February 2023].

Gokhale, P. Bhat, O. & Bhat, S. (2018) Introduction to IOT. International Advanced Research Journal in Science, Engineering and Technology 5(1): 41-44. DOI: 10.17148/IARJSET.2018.517

Huxley, K. (2020) 'Data Cleaning', in: Atkinson, P. (eds) SAGE Research Methods Foundations. London: SAGE Publications Ltd. Available from: https://dx.doi.org/10.4135/9781526421036842861 [Accessed 28 January 2023].

In Wong, H. (2023) Collaborative Learning Discussion 1: Peer Response Available from: https://www.my-course.co.uk/mod/forum/discuss.php?d=135855 [Accessed 11 February 2023].

Khalid, H. (2023) Collaborative Learning Discussion 1: Peer Response Available from: https://www.my-course.co.uk/mod/forum/discuss.php?d=135855 [Accessed 11 February 2023].

Marjani, M. et al. (2017) Big IoT Data Analytics: Architecture, Opportunities, and Open Research Challenges. IEEE Access 5: 5247-5261. DOI: 0.1109/ACCESS.2017.2689040

Ridzuan, F. Nazmee, W. M. & Zainon, W. (2019) A Review on Data Cleansing Methods for Big Data. Procedia Computer Science. 161: 731-738. DOI: https://doi.org/10.1016/j.procs.2019.11.177

Sarkar, D. T. Roychowdhury, S. (2001) Data Wrangling with Python. Packt Publishing. Available via Vitalsource Bookshelf. [Accessed 29 January 2023]

Sieminski, P. (2022) Collaborative Learning Discussion 1: Peer Response Available from: https://www.my-course.co.uk/mod/forum/discuss.php?d=135855 [Accessed 11 February 2023].

Tankard, C. (2012) Big data security. Network security. 7: 5–8. DOI: https://doi.org/10.1016/S1353-4858(12)70063-6

Yair, D. (2022) Your bible to structured vs. Unstructured Data, Bright Data. Available from: https://brightdata.com/blog/how-tos/structured-vs-unstructured data?kw=&cpn=14745430544&cam=aw_all_products-all_geos-search_dsa_blog-kw_en-desktop_blog-how-tos__612826796311&utm_term=&utm_campaign=all_products-all_geos-search_dsa_blog-kw_en-desktop&utm_source=adwords&utm_medium=ppc&utm_content=blog-how-tos&hsa_acc=1393175403&hsa_cam=14745430544&hsa_grp=136943771953&hsa_ad=612826796311&hsa_src=g&hsa_tgt=dsa-1649388330704&hsa_kw=&hsa_mt=&hsa_net=adwords&hsa_ver=3&cq_src=google_ads&cq_cmp=14745430544&cq_term=&cq_plac=&cq_net=g&cq_plt=gp&gclid=CjwKCAiAioifBhAXEiwApzCztq7RU_M_VKjByYivYWBtvbo0Ztza1ScLSkskwMSFwl2yj3BaAaIsLBoCqr4QAvD_BwE#key [Accessed: 11 February 2023].

Zainab, H. A. Hesham, A. A. & Mahmoud M. B. (2015) Internet of Things (IoT): definitions, challenges and recent research directions. International Journal of Computer Applications 128(1): 37-47.