The Importance of Big Data and Data Science in Understanding your Cash Cycle

Dr A. Leszczyszyn[1,*] **and Mrs R. Grindley**[1,**]

[1]7 Layer Solutions Limited, Cheltenham, GL50 3DT, United Kingdom
[*]Antin.Leszczyszyn@7layer.net
[**]Rebecca.Grindley@7Layer.net

ABSTRACT

This paper will discuss the importance of big data techniques in fully utilising the data obtained during the cash lifecycle. Data science methods used in conjunction with the big data techniques provide a full understanding of banknotes' lifecycles, stock levels, demand and quality.

Keywords: NoteChain, Big Data, Data Science, Cash lifecycle, Banknote Circulation

1 Introduction

A major role in Central Banks (CB) is to preserve the public confidence in banknotes. This confidence maintains a functional economy by upholding the fitness quality and authenticity of banknotes in circulation. Gathering information and gaining knowledge of banknotes' lifecycle is an important process in managing banknote demand and confidence levels.[1, 2]

Making policy decisions and banknote forecasts, CB need an understanding of how long banknotes can remain in circulation before they are no longer fit for purpose and need to be replaced with new banknotes. It is also useful to know whether: banknotes tend to return for destruction gradually over time, or a large proportion become unfit around a certain age. Data is at the heart of this understanding.

As discussed in 'What can be Gained with Serial Number Reading.'[3] High Speed Sorting (HSS) machines now gather substantial amounts of data, more than can be sensibly handled by traditional systems.

2 Data

The data collected can be used to:

- Make well informed decisions.
- Clearly define problems, which will lead to better decisions.
- Monitor decisions; were the right choices made?

- Forecast and predict.

However, once you've got your data, further questions arise:

- What do you do with it?
- What should you look for?
- What tools should you use?
- What hardware do you need?
- Who do you need?

These are all important questions answered below.

2.1 Big Data - Is it big enough?

With billions of bank notes in circulation it is easy to see that the data stored can get very large. Commercial Cash Handlers (CCH) and CBs use a variety of HSS machines, each producing a different set of data for the notes that they sort. The data output is dependent on the type of machine, its configuration and what it's sorting. By their very nature, HSS machines generate this data at high speeds. Traditional methods of data storage and analysis struggle to keep up with this increasing volume, variety and velocity. This is a common definition of big data.

2.2 Benefits of Big Data

Our paper 'What can be Gained with Serial Number Reading'[3] sets out some of the benefits of serial number reading which cannot be reasonably obtained without utilising big data techniques.

Even without adding serial number reading to the set of data obtained from HSS machines there are benefits to using big data techniques:

- Big data solutions will easily accept data from a variety of different sources, regardless of their format e.g. .csv, .xml, database.
- Many big data solutions provide inbuilt data security, an important consideration when implementing any solution.
- The distributed processing power of big data systems allows for complex data manipulation in a timely manner; keeping up with the growth of data provides up-to-date feedback on the cash lifecycle.
- Big data solutions are designed for commercially available off-the-shelf hardware. CBs can implement a solution on new or existing hardware.

2.3 Tackling big data

To obtain the benefits listed above from a big data solution, various components, and their combinations, need to be considered.

The components of a big data solution for CBs can be thought of as:

- Domain Knowledge; the knowledge of the cash cycle from within the CB
- Data; the data provided from your cash cycle
- Database Technology; for example, the Hadoop eco-system
- Mathematics and Statistics; analytical solutions derived from the data by statisticians and the industry experience of 7 Layer

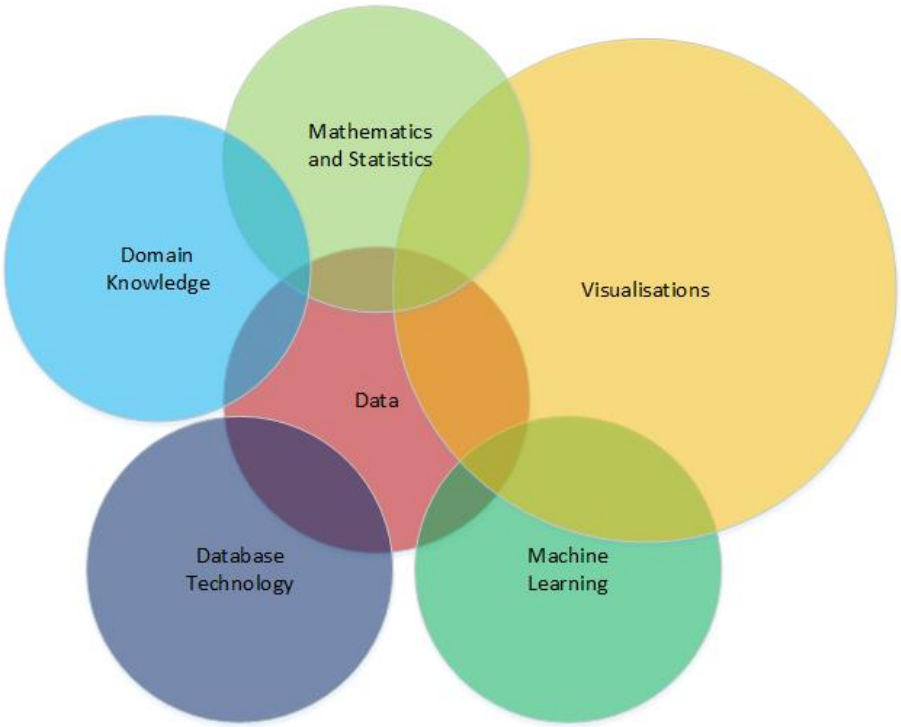

Figure 1. Data Science is the combination of mathematics, statistics, programming, database technology, problem solving, domain knowledge, capturing data in ingenious ways, the ability to look at things differently, and the activity of cleansing, preparing and aligning the data.

- Machine Learning; algorithms designed for a self-learning
- Visualisations; graphical representations of the data to provide clarity and ease of interpretation

These can be represented in the form of a Venn diagram, where Data is at the centre of a Data Science problem. Figure [1]

Every part of these components is improved by serial number. Connecting relationships are formed between each point of the data.

2.4 Domain Knowledge

This is where the CB's domain knowledge is applied. Data only makes sense within the correct context. The understanding of 7 Layer and CB domain experts, brought about by years of experience of the cash lifecycle, is incorporated in the solution.

2.5 Data

Data from various sources - HSS machines, contracts, printing etc becomes part of the data in the big data solution.

2.6 Database Technology

Alongside traditional databases there are now big data technologies that allow us to store larger amounts of data very easily.

As soon as it's available data is stored in the big data cluster. Here the data is cleansed to remove duplicates and malformed data before being normalised, bringing all the data to the same standard measure. Once at the same measure data is manipulated, directed by mathematical/statistical analysis of the dataset.

2.7 Mathematics and Statistics

As with the traditional methods statisticians, either from a CB or 7 Layer, apply analytic techniques to understand the data. With the availability of a larger set of data, including note serial numbers, statisticians will be able to focus their efforts on understanding past trends, current behaviour and predicting future events. A big data solution incorporating serial number tracking will allow this to occur with greater accuracy, deeper understanding and more rapid returns.

The quicker the data is processed, analysed and displayed to the user, the quicker conclusions can be drawn by the statisticians on the fitness quality and authenticity of the banknotes in circulation. Statisticians create questions to be answered and analyse the visualisations to inform the policy makers of the historic, current and potential future events in the cash lifecycle.

2.8 Machine Learning

Machine Learning is a newer tool in the analytics arsenal. Utilising algorithms designed to allow computers to learn and improve, machine learning aids in the discovery of more unknown unknowns in the data. Displaying these results to the end user allows them to be better informed about the historic, current and potential future events in the cash lifecycle and to contemplate questions that may not have otherwise occurred to them.

2.9 Visualisations

Visualisations allow the end user to make sense of the reams of data produced by the various sources. Presented in the form of charts and graphs they are grouped to provide answers to many of the questions faced by today's CBs. Visualisations have been in use for a long time, from individual hand drawn graphs painstakingly created from small sets of data to graphs produced from tables of data stored in Microsoft Excel, big data technologies allow instant insights to be drawn from much larger sources of data than previously.

3 Conclusion

The growing data available in the cash lifecycle needs to be harnessed in order to review previous policies, inform new policy decisions and generate banknote forecasts. Big data techniques are the ideal way of accomplishing this.

7 Layer understands the need for big data and how it applies to the cash cycle. We understand how to create solutions using many methods with big data.

NoteChain® is a big data solution using all the above methods to construct an easy to use solution for your cash cycle.

References

1. Rush, A. (2015), *The Life of Australian Banknotes,* Reserve Bank of Australia, Research Discussion Paper, 10.
2. Deinhammer, H. & Ladi, A. (2017), *Modelling euro banknote quality in circulation Feedback,* European Central Bank, Occasional Paper Series, 204.

3. Leszczyszyn, A.M. & Grindley, R.M. (2018), *What can be Gained with Serial Number Reading,* 7 Layer Solutions Ltd, White Paper.