



THE PROPERTY C H R O N I C L E[®]

ESTD. 2017

Real estate, alternative real assets and other diversions

STEPHEN RYAN

UNSTRUCTURED DATA IS YOUR FRIEND

March 30, 2020



THE ANALYST

Recent Posts

[HOW TO GET IT RIGHT WITH REITS](#)

[HOW IT WAS](#)

[HOW I DISCOVERED THE BEAUTY OF PRIVATE DEBT](#)

[HOW TO CATEGORISE STOCKS AS QUALITY DEFENSIVE AND/OR](#)

Unstructured data is your friend

Data is either structured or unstructured. Structured data is neatly arranged in rows and columns, like a spreadsheet. Unstructured data is messy and disorganised. It shambles towards us as PDFs, Word documents, presentations, open-ended survey responses, webpages, blog posts, social media sites, audio files, images and emails.

Both types of data can support decision-making. For example, the free-form responses in a tenant survey enrich the data gathered from structured questions, and landlords benefit from both. In real estate more attention is paid to structured data, where the analytical processes and technology are mature. But unstructured data is also valuable. And there is far more of it.

A pair of wranglers

On this side of the Atlantic to wrangle means to argue or bicker. On the American side it means to round up cattle, horses, or other livestock. More recently, it has acquired a new meaning - wrangling data has come to mean the process of cleaning and unifying messy data sets for easier access and analysis. The messiest data is likely to be the unstructured data. A data wrangler is someone who integrates information from multiple sources and transforms it into something more useful.

If your firm's real estate data comes from disparate sources, or is stored on a scattering of legacy systems, you might need to try on some wranglers.

Initial trawl

When working on client research projects at Didobi, we often start by reviewing the existing publicly available data. This means marshalling a vast amount of data from different places, and much of the data is unstructured text data (for example, articles from the trade press, FCA consultation papers, research from trade associations such as AREF or IPF, legislation and academic papers).

Locating and reading the relevant material is time-consuming but essential. A skim-read will not suffice because even the most attentive reader can miss something. Sometimes the most revealing nugget of information is hidden deep in an appendix.

Needle in a haystack

We are not looking for a single needle in a haystack, we want all the needles. That means identifying the most promising data sources, and then considering every single line of text in those sources. Doing this properly requires either an army of assistants or a computer. Missing something is less likely when we get the computer to read and summarise for us.

My computer reads and summarises files for me. Does yours?

A computer can read through a cleaned-up file and summarise it in less than one second using natural language processing (NLP). NLP is a branch of artificial intelligence that helps computers understand and manipulate human language. It can break down language into shorter, elemental pieces. This helps us to retrieve the essential information from text-based sources.

In our case, we get the computer to assemble an extractive summary - a subset of words and sentences containing the key points, using only words found in the original data. The summary is extracted by applying rules to rank words and phrases. Armed with a summary of each potentially interesting file, we can identify those which merit closer (i.e. human) scrutiny.

Flying on one wing

There is plenty of unstructured data for the real estate researcher to investigate, and the volume of it is growing every year. As buildings and cities become more interconnected, the volume will explode. Our industry is surrounded by unstructured data but is not yet extracting the maximum value from it. Combining structured and unstructured data will lead to better decisions and better outcomes. Do not waste your valuable data. It is time to start wrangling.