



## Mission: Possible

Ashley Keil, ibml's VP sales, EMEA/APAC discusses how BPOs and enterprises can close the gap to getting 100% data accuracy as part of their capture processes and business workflows

If you thought the missions that 'Impossible Missions Force' agent Ethan Hunt undertakes in the popular action spy series are tough - or even impossible! - then spare a thought for CIOs, IT and information management professionals tasked with looking after company data. Their mission to manage data is getting considerably more difficult as data arrives from more sources, in more formats, of varied quantity and in greater quantities than ever before.

This is reflected in the latest IDC global research, published in May this year, which highlights the continued growth of data. They say that over 59 zettabytes will be created, captured, copied, and consumed in the world in 2020, with the amount produced over the next three years predicted to exceed the amount created over the past 30 years. That's mind-boggling.

Yet according to Gartner, 40% of an enterprise's data is inaccurate, missing or

incomplete at any given moment in time; 13% go so far as to rate their data quality as poor; with only 47% of organisations surveyed saying they have high quality data.

The implications are significant. Yes, it might be a cliché, but data is most definitely the lifeblood of any business and government organisation as it feeds back-end processes, powers decisions and fuels profits.

No matter your industry therefore, obviously inaccurate data is detrimental and no one wants it in their systems. Get it wrong and you're into a whole world of pain. Bad data erodes operational efficiency, slows down decision making, stunts ROI, makes delivering SLAs tricky, adds commercial risk, delivers poor customer experience and damages relationships. Ultimately it's bad for your bottom line too, with data governance very much part of GDPR rules and the associated penalties and fines.

### CURRENT DATA PRACTICES ARE GOOD, BUT...

But it's not all doom and gloom. Many organisations and their BPO service partners have made considerable headway automating data capture processes successfully, investing significantly in best-of-breed intelligent capture technology which integrates easily into line of business systems because of the use of open APIs. This helps expedite processing the tsunami of information coming in whether it's extracted from postal mail, email, fax, images from smartphones or other sources.

Artificial Intelligence and machine learning platforms today perform complex data capture with minimal operator invention. We're talking accuracy rates of anywhere between 80 and 95%. The variation comes when you have to deal with, for example, crumpled or torn paper, text where a highlighter pen has been used or illegible handwriting on a form.

"MANY ORGANISATIONS AND THEIR BPO SERVICE PARTNERS HAVE MADE CONSIDERABLE HEADWAY AUTOMATING DATA CAPTURE PROCESSES SUCCESSFULLY, INVESTING SIGNIFICANTLY IN BEST-OF-BREED INTELLIGENT CAPTURE TECHNOLOGY WHICH INTEGRATES EASILY INTO LINE OF BUSINESS SYSTEMS BECAUSE OF THE USE OF OPEN APIS. THIS HELPS EXPEDITE PROCESSING THE TSUNAMI OF INFORMATION COMING IN WHETHER IT'S EXTRACTED FROM POSTAL MAIL, EMAIL, FAX, IMAGES FROM SMARTPHONES OR OTHER SOURCES."

It's just more challenging for the recognition engines to extract and convert this kind of information into ASCII files so that the data can be ingested into downstream business processes.

To boost accuracy rates, barcode technology has been used with much success. But they are not a panacea and only work with a small percentage of data capture situations. Amazingly, in the quest for perfect data, some organisations have resorted to employing staff to manually re-key information or by relying on operators to review capture results for each document to ensure accuracy. These approaches to eliminating data errors are costly, time consuming and far from foolproof.

So, what are the options if accuracy rates of 80, 85, 90, 95% or whatever aren't good enough in a commercial situation? How can the 'last mile' - so to speak - of data capture be improved to get to the nirvana of 100% without the considerable expense of adding more headcount?

#### DATA PERFECTION IS ATTAINABLE

The answer lies in using a multifaceted approach optimising a mix of four main components:

1. Best of breed capture technologies;
2. Rules-driven capture and validation;
3. AI-driven matching;
4. Human and AI-powered triple data entry.

The use of capture technologies will be familiar to DM readers. What might not be quite so well-known is just how fast and powerful some of the hardware is today. High performance intelligent scanners - like our FUSION platform - now process volumes up to 730 A4 pages per

minute, with FUSION designed to be ergonomic for bureau staff to use whilst offering low operational expense in terms of maintenance.

These scanners come with real-time, in-line intelligence that helps understand documents, extracting data early in the process so as to minimise errors downstream. Importantly, business rules can be set to capture and validate field-level meta data. So, for example, the scanner will review whether an application form has a signature or if exams scripts have the right numbers of pages and are in the correct order. Remedial action can be programmed in if they don't. To repeat, this occurs as it happens in real time as documents are literally in motion on the scanner.

In addition, AI-driven matching solutions are available - integrating with the scanner or independent of it - to enable the cross-referencing and matching of multiple incomplete or incorrect data fields against master database sources so that errors can be flagged and dealt with immediately.

This means that a number of partial metadata captures, which are inaccurate in their own right, can be pieced together and combined to correct and validate the information being processed before it is accepted into a business system. A very simple example would be scanning mail. An envelope might be muddy or damaged, obscuring bits of the name, address, postcode or all three. By assessing all the fields and the text and then cross referencing this extraction in a master database - which might hold millions of customer records - the AI solution can bring these partial 'reads' together to get a qualified and accurate result. Complex algorithms are used to do this, taking just milliseconds.

#### GETTING HELP FROM THE CROWD

The fourth way to achieve clean data is to use a scalable automated crowdsourcing approach to do what's called triple data entry. This pretty much guarantees data accuracy. It's ideal for a range of applications like forms and loans processing, prescription management, mail room, customer onboarding and so on.

Crowdsourcing pushes snippets of the same information to online data entry clerks based globally who are connected to a management platform via the Internet. Two people then check the same snippets of unmatched or poor quality data from an image before entering it into a system. If there's a mismatch between what the two individuals then input, it goes to a third person for exception handling which solves the issue of manual errors creeping in. This is how 100% accuracy rates are achieved.

Crowdsourcing data checking is ideal where intelligent word or character recognition technologies - ICR and IWR - have struggled to recognise handwriting in a field and more validation is required. Self-evidently working with a specialist crowdsourcing partner is a fraction of the cost compared to physically employing staff with all the associated expenses of salary, pension, office space, desktops and so on. The data entry operators get paid per key or entry stroke based on the platform they are signed up with.

Data's exponential growth has created opportunities to leverage it in new ways for better business outcomes. Accuracy is therefore key. Crowdsourcing is a relatively new area in the information and document management industry. This kind of data validation approach is cost effective, fast, secure and works reliably which leads me on to say: "Your mission, should you choose to accept it, is to give it a go".

**More info: [go.ibm.com](http://go.ibm.com)**