Are you the Master of your Data?

I remember talking with an industry peer about how to make data quality issues clear to people who are not data specialists. She told me about her initiative that she called "12 Christines".

What she did was to find all the different instances of herself as a customer in her company. Each different product or service that she'd purchased from her own company had created another "Christine" record without checking to see if she already existed. She was able to find 12 different customer records across a range of systems that all related to the one person, herself. Taking this information in front of her management provided a clear example of the pain of poor Master Data management. Showing the actual tangible impact of poor data quality and poor master data management was very effective in bringing the costs home to the management and arranging funding.

I was reminded of this when I was watching the recent Australian Federal Election. I found that my preferred newspaper was being very slow at updating the seat-by-seat election results, so I went directly to the Australian Electoral Commission's website, and I spotted a Master Data Management issue in their data. Have a look at the screenshot below and see if you can spot the problem?

Divisions where the incumbent is trailing for all states and territories

Division ⇕	State ⇕	TCP counted % ⇕	Incumbent ⇕	Led by ⇕
Swan	WA	97.83	Liberal	Australian Labor Party
Pearce	WA	98.38	Liberal	Australian Labor Party
Chisholm	VIC	98.41	Liberal	Australian Labor Party
Grey	SA	6.26	Liberal	Australian Labor Party
Wentworth	NSW	99.84	Liberal	Independent
Hasluck	WA	100.00	Liberal	Australian Labor Party
Higgins	VIC	78.85	Liberal	Australian Labor Party
Reid	NSW	100.00	Liberal	Labor
Kooyong	VIC	75.57	Liberal	Independent
Tangney	WA	84.18	Liberal	Australian Labor Party
Goldstein	VIC	100.00	Liberal	Independent
North Sydney	NSW	99.70	Liberal	Independent
Boothby	SA	100.00	Liberal	Australian Labor Party
Mackellar	NSW	100.00	Liberal	Independent
Robertson	NSW	100.00	Liberal	Labor

Did you spot that the Australian Labor Party has two different names in the right-hand column? In some cases, its referred to as Labour, whilst in other cases it is referred to by the full name of Australian Labor Party. This poor Master Data Management outcome will make analysing AEC data significantly more difficult for analysts, media and political parties. These two different names will show as if they are two different parties, which will result in an apparent worse showing for what is one party.

Managing master data effectively requires several different disciplines working together to achieve. These are:

- Effective training of all data entry staff
- KPIs that motivate speed AND quality of data entry, not just speed
- A smart central repository that acts as the master source for whatever master data dimension(s) that need managing
- An effective ability to query in real time against the master source.
- Front-line systems (web, call centre, data entry etc.) that can query for existing data against the master source, and re-use that data if an existing master record is found.

The smart central repository needs fuzzy matching capabilities such as phonetic matches, point-based percentage match processes and creation of golden master records based on variable inputs.

Are you well placed to master your data, or is your data mastering you, as it appears to be at the AEC?

If your business would like help with this, please contact a Fujitsu Data & AI specialist now.

Contact

Fujitsu Data & AI
+61 3 9924 3000