

## Batch cleansing of addresses

---

With the best will in the world, a batch cleanse, whilst quicker and more cost effective is unlikely to produce better results than someone sitting down and going through each record, one by one and fixing them up. But for most companies, manual cleansing is not a sensible option with regard to time or money.

Batch cleansing has to provide a solution for the majority of issues found within a file of addresses, with a narrow point of reference – the Postal Address File (PAF) and in many cases a series of lookup tables.

Acxiom has worked with InfoSoft to provide their clients access to cost effective batch cleansing services via the Acxiom Data Portal (ADP). In order to try to keep this as simple and as effective as possible, a set of templates have been set up on ADP for InfoSoft clients. The templates are as follows:

### Job Specification Templates

These pre-select the work that is to be done on the file e.g.:

- Address Hygiene only
- Address Hygiene and NZ COA
- Opticx (Audit report)

### Input Template

These map in the various address lines so the parser knows which fields it should be looking at, and which ones it can ignore

### Output Template

This tells the software where to place the address elements and in what format the file should be output

Accessing Address Hygiene services via the Bureau service gives a company access to more tailored cleansing. Issues that are specific to your data set can be allowed for and processing tuned accordingly. Accordingly, the price will reflect the time and effort required to process this data.

When cleansing via an automated batch cleanse, such as the ADP option, the process has to be set up to handle the majority of issues. If we were to tune the cleansing to allow for specific issues within particular datasets, it may well be perfect for that one set of data, but could potentially cause catastrophic results for another set of data.

### So how does batch cleansing work?

Generally the first part of a batch cleanse will involve a parsing process. This is where the software tries to identify the various elements of the address e.g. Flat type, Flat number, Building name, Street number, Street Name etc.

Most of this information can be identified via referencing the PAF or lookup tables. E.g. Company names usually have an identifier such as Ltd, Pty etc., and many building names are already in the PAF.

It is also generally accepted that Company Name info, Building names, Flat info etc will be located in the first couple of address lines.

Where this can become a problem is when it finds information that it is not expecting within the address lines, and/or information where it wouldn't expect to find in those address lines it e.g Company names with no identifier, Individual names, Job Titles, Building names that aren't in PAF etc.

If we were to tell the software to exclude non-PAF information (i.e. if it doesn't appear in PAF remove it) this information wouldn't cause many problems and the address would have a better chance of being verified, however information that may be valuable to your business would be dropped.

The parser will do it's best to try to relocate this information to relevant fields, but this will depend on the information and how and where it is presented. If it can't identify the information, it will try to output it to a field called "ADP\_ADDRESS\_OTHER", so you can view it and confirm whether this is information that you would like to keep or discard.

The following examples have been taken from InfoSoft client datasets, however names and some addresses have been changed – the placing of the elements are as presented.

#### Example 1:

Full Name including title in Address Line 3. Not only does the parser not expect to find this information in this field, there are some building names in PAF that start with "Rev" and "Reverend". There is a very good chance that this information will be output in the "FLOOR\_BUILDING" field.

ADD_PART1	ADD_PART2	ADD_PART3	Suburb	City	State	PostCode
		Pastor John Smith	13 Hughes Street	Waikanae Beach	Waikanae	5036

ADD_PART1	ADD_PART2	ADD_PART3	Suburb	City	State	PostCode
		Rev John Smith	13 Hughes Street	Waikanae Beach	Waikanae	5036

**Example 2:**

There are no “Corner” addresses in PAF. In normal circumstances the Parser will try to identify a single street address (Daniell Street) and would output Cnr Constable to “ADP\_ADDRESS\_OTHER”. Because we have found this is quite common in the InfoSoft client data we have seen, we have written some code that looks for this kind of information and outputs it into the “ADDRESS\_POSTAL” field.

ADD_PART1	ADD_PART2	ADD_PART3	Suburb	City	State	PostCode
Cnr Constable & Daniell Streets			Newtown	Wellington		6021

**Other things that can affect the quality of your address cleanse**

**Splitting information across address lines.** The parser will try to identify address information if it is present, but this information is likely to be output in “ADP\_ADDRESS\_OTHER”

FIRSTNAME	LASTNAME	ADD_PART1	ADD_PART2	ADD_PART3	SUBURB	ADP_TOWN_CITY	POSTCODE
K M	Baxter		and Mr D Kent	5 Corrunga Avenue	Newtown	Wellington	6002

**Presentation of Street Ranges**

ADD_PART1	ADD_PART2	ADD_PART3	SUBURB	ADP_TOWN_CITY	POSTCODE
	12 – 14 Sunrise Boulevard	Longview Rest Home	Redwood	Wellington	6006

New Zealand Post no longer accept street ranges as part of the address and there are no indicators on PAF to advise if an address may be part of a street range. Previously New Zealand Post used 8/10 to indicate a range and 8-10 to indicate flat/unit. Now they use 8/10 to indicate flat/unit, so the data we receive can be well and truly mixed up.

The only way the parser can handle this issue is to err on the side of caution. The only time it even considers that something may be a range is if there is already a flat unit number in the address:

- 8-10 Smith Street is seen as Flat 8, Number 10
- 1/8-10 Smith Street is seen as Flat 1, Number 8 to Number 10

**Presentation of Flat information**

There are a number of issues surrounding the presentation of Flat/Townhouse/Unit/Apartment/Villa information. Some of your donors will be particularly sensitive about this, unfortunately we only have the PAF to reference for this information.

If your record say Apartment 1, 13 Hughes Street, and in the PAF there is only a Flat 1, 13 Hughes Street, it is not always safe to assume that this is the same property, particularly now that inner city apartments have become more prevalent.

In some apartment buildings we find Flat, Apartment and Suite addresses, along with a floor number and depending upon the presentation of that combination it can be very confusing for the parser.

Here are some real examples of the presentation of this kind of information:

ADD_PART1	ADD_PART2	ADD_PART3	SUBURB	ADP_TOWN_CITY	POSTCODE
	1.14 Queens Wharf Apartments	15 Jervois Quay		Wellington	6001

ADD_PART1	ADD_PART2	ADD_PART3	SUBURB	ADP_TOWN_CITY	POSTCODE
	No 2 Tower Level 7 Shortland Centre	55 Shortland Street		Auckland	1001

ADD_PART1	ADD_PART2	ADD_PART3	SUBURB	ADP_TOWN_CITY	POSTCODE
	10 B – 20 Oriental Terrace	Jerningham Apartments	Oriental Bay	Wellington	6011

The correct output for the address above is as follows:

UNIT_FLAT	FLOOR_BUILDING	ADDRESS_POSTAL	SUBURB_LOCALITY_RD	TOWN_CITY	POSTCODE
Suite B	Floor 10 Jerningham Apartments	20 Oriental Terrace		Wellington	6011

### Casing

The PAF is supplied in Upper Case. Whilst New Zealand Post allows you to choose the casing you, their preferred format is Proper Case, with Country in UPPER CASE. Unfortunately, without the presence of Casing Tables it makes it difficult to correctly case elements such as building names e.g. AMP House. You will find that your data is returned in Proper Case e.g. Amp House.

### FAQS

Questions are shown in italics.

*Can you please clarify what the different groups of data refer to that are now part of our Favorites for manual checking after your initial data cleanse?*

- ***DPID assigned at primary point level***

A DPID has been given but the address needs checking

This is where you may have, for example, Flat 1, 107 Church Road and PAF has 107 Church Road and 107A Church Road but no Flat 1, 107 Church Road. The DPID for 107 Church Road has been applied, as this is the “Primary Point” however you may need to check with your donor as we can’t assume that Flat 1 is in fact 107A.

- ***New postcode given only***

System knows the postcode but does not know the address

In these cases we have tried to append an up to date postcode based on the information in the address, however there is not enough information, or the information that is there is conflicting so a fully verified address could not be output.

- ***Additional data found in address***

Non PAF data is in the address

The address has been verified 100% to PAF however there was information in the supplied address (and in most cases we have tried to output that in ADP\_ADDRESS\_OTHER for you to view) that does not appear in PAF. For

example, you may have had both a PO Box and a Street address in the same record. We always try to output the PO Box as the verified address and output the Street address in the OTHER field. You should always check these records as there may be information that you do not want to lose but that doesn't belong in the address itself.

- **Other reason**

Data cannot be matched with PAF

- **Delivery type not found**

Cant recognize PO Box Number etc

These are generally where a PO Box number doesn't fit with either the Box Lobby (Suburb) or Postcode that has been supplied.

- **Number not matched for delivery type**

Address may contain both postal and delivery data. Can also be where you have supplied a Private Bag with no number, or a Private Bag when it should be a PO Box.

*Our main question at this point is can we identify these in away that relates to the type tags we have on our database and/or most importantly through the financial area showing us if they have given us money at any time. This would make it easier to see which Contacts have more of a priority to check details and include in any upcoming mail outs and then include in our next data cleanse so we can obtain our Statement of Accuracy.*

Answer: Acxiom provide us the data – it does not match system data in anyway

*For those that have never supported financially is it possible to group them together separately and ideally Archive and don't include in our Statement of Accuracy at this stage?*

You can submit any data you want an SOA for.

*As you know we have many addresses that have not been active for some time, this is making it difficult to check things manually especially when they don't appear in the Whitepages which at this stage is my main avenue for verifying details.*

*Can you also clarify how addresses are handled with the datacleanse that are part of Resthomes or similar that may have a Unit, Flat or Apartment, sometimes an individual's address is also in a similar layout in that they show a Flat or Unit number but the NZ Post street address is only showing the main street address. I have found out that I can override this by verifying with the main street address and then adding the Flat number and not verifying on the second occasion, of course this means that DPID number is no longer showing. Will this affect any check / cleanse you then do for us?*

On the Fundraiser Today page you will see a link to a document that helps with searches

*What happens to those addresses that at this stage we don't need to include for NZ Post ie.*

- *Un-numbered rural addresses*
- *Private Bags with no number*
- *Addresses maintained by other postal operators eg. DX, Mail Boxes etc*
- *Community Mail Box*
- *Poste Rest Ante*
- *Overseas/International*
- *Counter Delivery*

All of the above are ignored for SOA, provided that they are presented correctly so the parser can identify them. E.g for a rural delivery address to be ignored, it has to have no street name or no street number and no street name, it must have an RD number, a Mail Town and an up to date Postcode.

e.g:

1 Horseshoe Bush Road, RD 4, Albany, 0794 – this record will be INVALID as it does not appear on PAF

Horseshoe Bush Road, RD, Dairy Flat – this record will be INVALID as it does not have an RD number, the Mail Town is incorrect and it doesn't have a Postcode

Horseshoe Bush Road, RD 4, Albany, 0794 – this record will be IGNORED for SOA

RD 4, Albany, 0794 – this record will also be IGNORED for SOA

*I have noticed that for some reason some of the addresses that are PO Box addresses are showing on address line 2, with suburb on address line 3 is there anyway this can be picked up in the cleanse so that they are in the correct area.*

Most will be in correct area, some that contain other data may not so you need to do this manually.