Data Cleaning

The purpose of this document is to help you avoid the costs of poor data quality.

We look at how to find data issues such as duplicate entries and avoid them happening where possible. We also look at efficient ways of fixing the problem.

The data records used in this document are about imaginary people “demo” data. Any similarities with real people are un intended

Table of Contents

1 Prevention and diagnosis 3

1.1 Finding companies / individuals before adding a duplicate 4

1.1.1 The top left search field 4

1.1.2 The multi search 4

1.1.3 A Report single field Search 5

1.1.4 Company tab and the Contact tab 5

1.1.5 Maps 5

1.1.6 Naming conventions 6

1.1.7 Authority to view (What if the company/contact is already there but you simply cannot see them?) 6

1.1.8 Avoiding duplicates at source using automatic tools 7

1.1.9 Telephone call screens with no or multiple caller records 7

1.1.10 Maps 9

1.2 Add companies and individuals (including profiling, etc.) 10

1.2.1 Type in 10

1.2.2 Delegating data entry 10

1.2.3 Online forms and dedupe rules 11

1.2.4 Importer. 11

2 Cure: Tidying up data 11

2.1 Data maintenance 11

2.2 Deletion permissions and caveats 11

2.3 Deduplication process 11

2.4 Example Process 12

2.4.1 Transfer any jobs 13

2.4.2 Transferring client Contacts from one account to another 13

2.5 Allowing deletion for a user 15

2.6 Transferring jobs/tickets/requests from one contact to another 18

2.7 Special cases: 21

2.7.1 Email bounces 21

2.7.2 Incoming calls 21

2.7.3 Staff database clean ups 21

2.7.4 Marketing annual opt in/ out surveys. 21

2.8 Other notes: third party data cleaning services 21
2.9 GDPR 22
2.9.1 right to be forgotten 22
2.9.2 right to correct data held 22
2.9.3 right to see data held 22

1 Prevention and diagnosis

There is a cost to incomplete, late or generally inaccurate data and it can vary with the problem. Here are just a few common examples. There are others which have cost organisations millions of pounds.

Data quality issue	Potential consequences
Wrong or missing telephone number	Time wasted by many colleagues looking for the number or disturbing other colleagues to ask.
Name spelled wrong	Client feels unvalued and may ultimately be insulted. Shows lack of care which may be conflicted at a bad time for your service level.
Work not logged correctly	MIS wrong – potentially poor decision making. Work missed or not prioritised appropriately. Searches on previous work for established solutions is ineffective and research and investigation must be repeated.
Email not cleaned	MIS works out wrong. Unsubscribe on one contact leaves duplicate contact subscribed to that newsletter. Preferences stated once are ignored in other contact. Knowledgebase about that person fragmented. GDPR compliance difficulties

Obviously, in high volume businesses, this can cost millions. However, even in the smallest business damage can be done sending a single inappropriate email.

The best way to have clean data is undoubtedly to have processes in place o keep it clean in the first place! This means enforcing organisation wide discipline.

- Always add contacts, addresses and contact details and company names accurately – even if they are unclear when you first come across them.
- Use automation. For example, the telephone integration will throw up client names if their incoming telephone numbers are correct and will bring up blanks if they are not recorded against clients in your system.
- Comply with your organisation's naming standards. (Please see below.
- Always add notes during your conversations. Always attach files straight away.

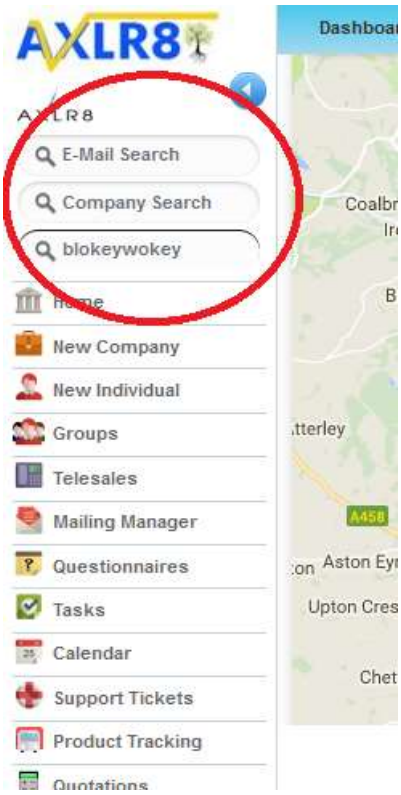
1.1 Finding companies / individuals <u>before</u> adding a duplicate

This is how to find companies as clients call or email in or as you are dealing with normal business.

There are several ways to find if a company is already on the system:

1.1.1 The top left search field

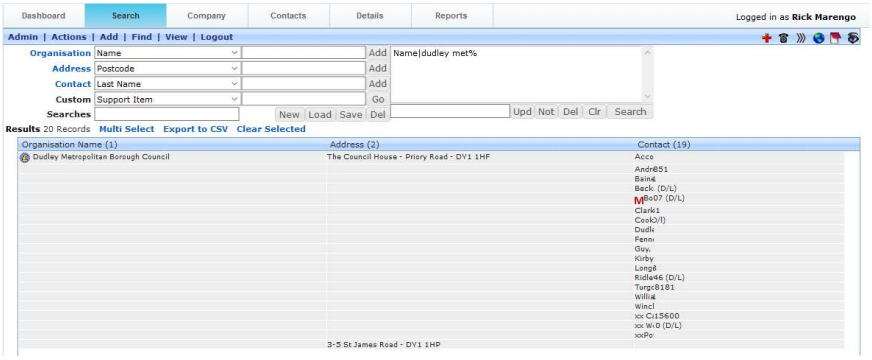
The simplest way to search is in the top left Search field.



As you can see, you may look for an Organisation (company, etc.) or you can search for a Contact by their email or last name.

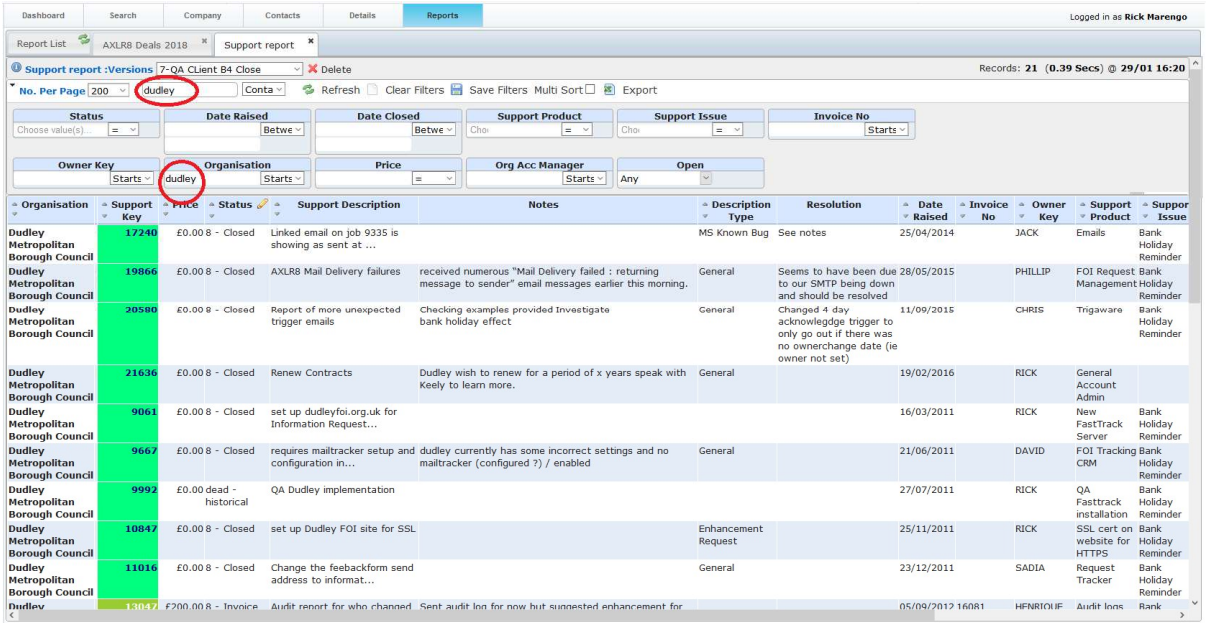
1.1.2 The multi search

In the Multi Search, you can, for example, search by a company and other combinations of parameters. In this case, I have simply searched for any organisation starting with "Dudley Met".



1.1.3 A Report single field Search

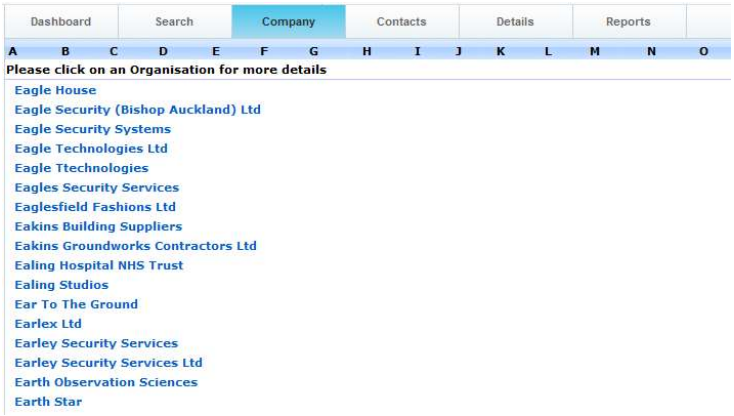
The reports function is a very powerful search and list tool for managing data quality, spotting outliers and many other purposes>



1.1.4 Company tab and the Contact tab

In the Companies Tab, you can see companies as shown in this demo screenshot. Just click on the first letter to see the list.

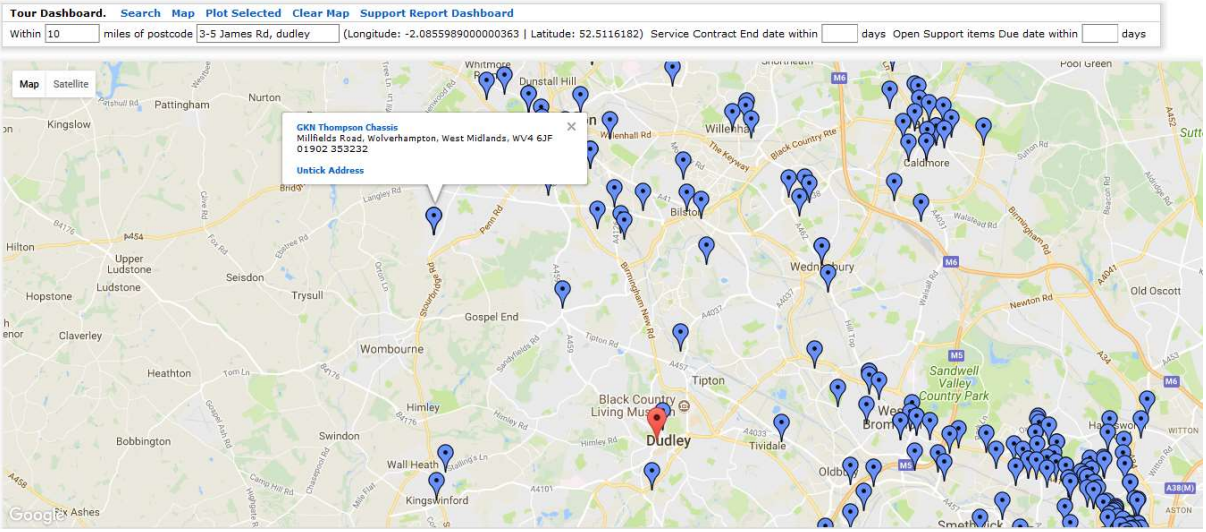
Because they are in alpha order, it is sometimes easy to see duplicates.



Likewise in the Contacts tab, you get alphabetical contact lists.

1.1.5 Maps

Maps are a particularly good way of assessing whether your address data is correct. It is also an effective trip planning economic tool that pays you back for the time you invested entering accurate address data.



1.1.6 Naming conventions

Nobody wants to look for both BBC and B.B.C. and British Broadcasting Corporation so have a company standard on how you deal with these. Likewise, if you are looking for "Hewlett Packard" and Someone typed it in as "Hulet Packard", it will probably not be found and someone will then (a) not have access to the account history and (b) compound the issue by entering a duplicate.

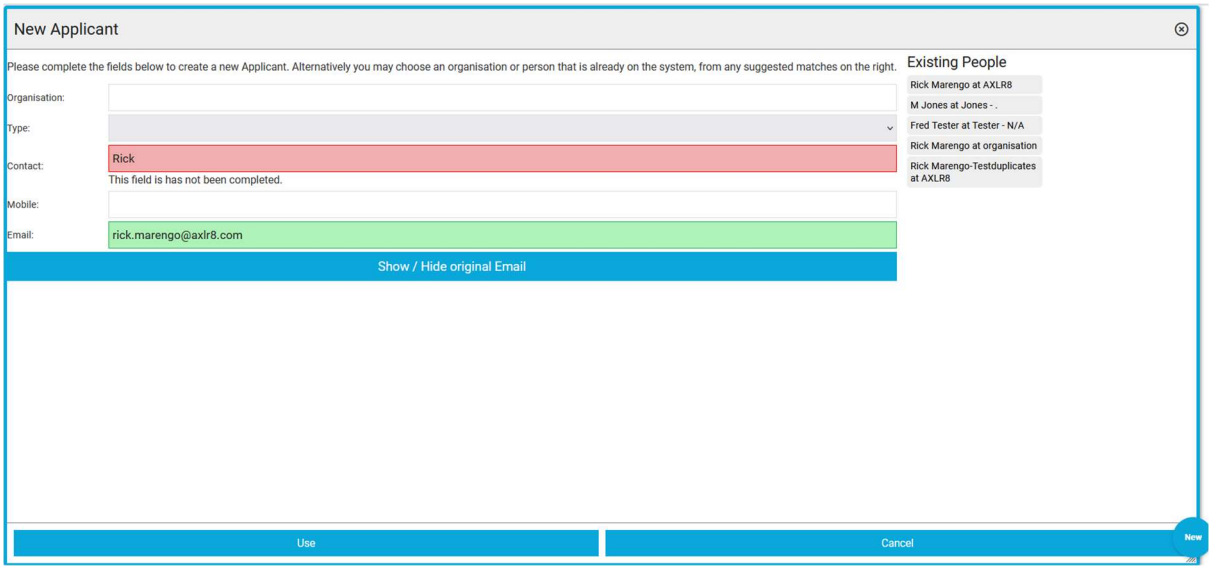
Make sure your organisation agrees and sticks to naming conventions.

1.1.7 Authority to view (What if the company/contact is already there but you simply cannot see them?)

Most modern systems for commercial and public sector have some form or de-duplication mechanisms built in by AXLR8.

1.1.7.1 Public Sector

In the public sector a new email information request will "find" a previous instance where that email came in. Sometimes, however, the email changes from request to request such as when the email is generated by services like "WhatDoTheyKnow" or from a reporter when they write for different periodicals/ papers, blogs, etc. In the screenshot below, we see a new applicant and there are alternative names that the system is suggesting may be duplicates on the Top RHS.



1.1.7.2 Sales teams entering new prospects

When a sales person adds a new prospect., many of our systems will have ways of searching on both B2C and B2B contacts and organisations as they are being typed into the system. Various options are offered.

In addition to human decision making, many of our financial clients' sales people are asked to add a company registration number early in the process. This not only brings back accurate data from Companies House during the entry process but also flags up if the company is already on the system to avoid patch disputes.

1.1.8 Avoiding duplicates at source using automatic tools

1.1.8.1 Staffing

In ATS (applicant Tracking systems) the application form will inform the candidate who attempts to apply again that they should contact the office. The behaviour and parameters vary from client to client but options include deduping on email address (most popular), mobile number & NI number.

The job boards sometimes create new emails such as joe.bloggs3456789@indeed.com which different for each application. So ask AXLR8 about our Indeed API implementations to reduce or prevent this. You may still not find the individual even it they exist in the system. If you have hundreds of applicants coming through each day, you can see how the problem can cascade.

1.1.8.2 Web forms

Several commercial clients have web forms which take new client enquiries and add them to the system. These deduplicate typically on email address. This is effective in a B2B environment usually.

1.1.9 Telephone call screens with no or multiple caller records

When this call comes in, I can see all the details about that client.

It is the switchboard for Dudley Metropolitan Borough Council. As someone from that client calls in, they would normally use their direct line or perhaps a mobile so I can see exactly who it is and have immediate access to their support tickets. If their number is not there, our service will be a tad slower.

If they are a client, and if the user has the authority, then they may see all the contacts and all the invoices, etc.

However, in this case (screenshot below, cellphone number removed). I can see there is a duplicate and deal with the matter immediately. In this case, we can see from the audit logs that the company was set up by the same user on AXLR8's internal CRM within the same hour so they probably forgot they had done it – perhaps interrupted by another call.

The process for removing duplicates is described below.

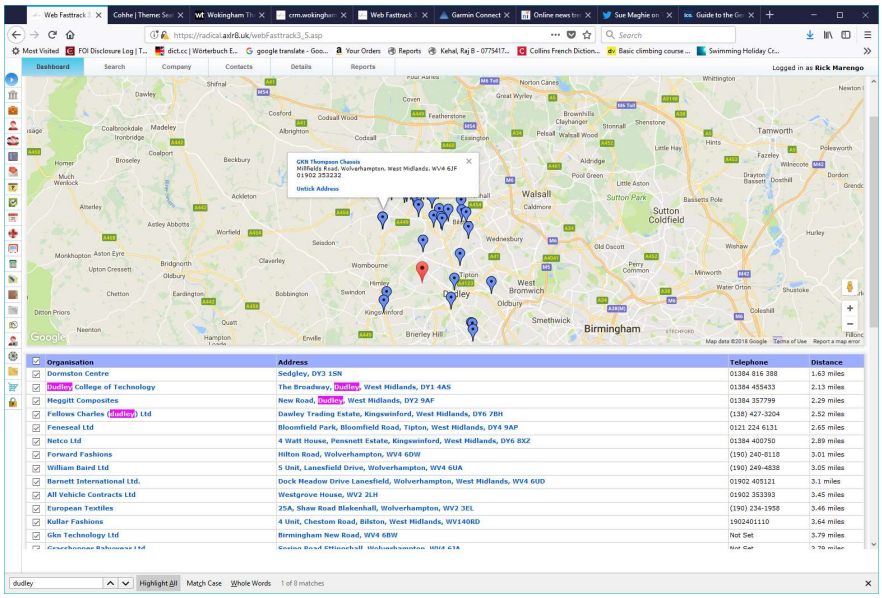
Remember, some similarly named duplicates are supposed to be in there. Check carefully with your data analyst/ boss/ colleagues before. It is always better to flag things for deletion for a consultation period in case a colleague needs that data. In an AXLR8 client I spoke with yesterday, they have more than a dozen organisations which start "Alcatel…….". They are each very different legal and commercial entities as far as my client is concerned. They have different sales people and commercial terms.

Also, don't forget that an incoming number may provide valid multiple organisations. Many entrepreneurs and "portfolio" players as well as consultants could call in using their mobile and multiple organisations can pop up. As I wrote this, I had a call with an entrepreneur from Dubai and his 5 companies came up on the list each had several contacts in different offices with very few in common.

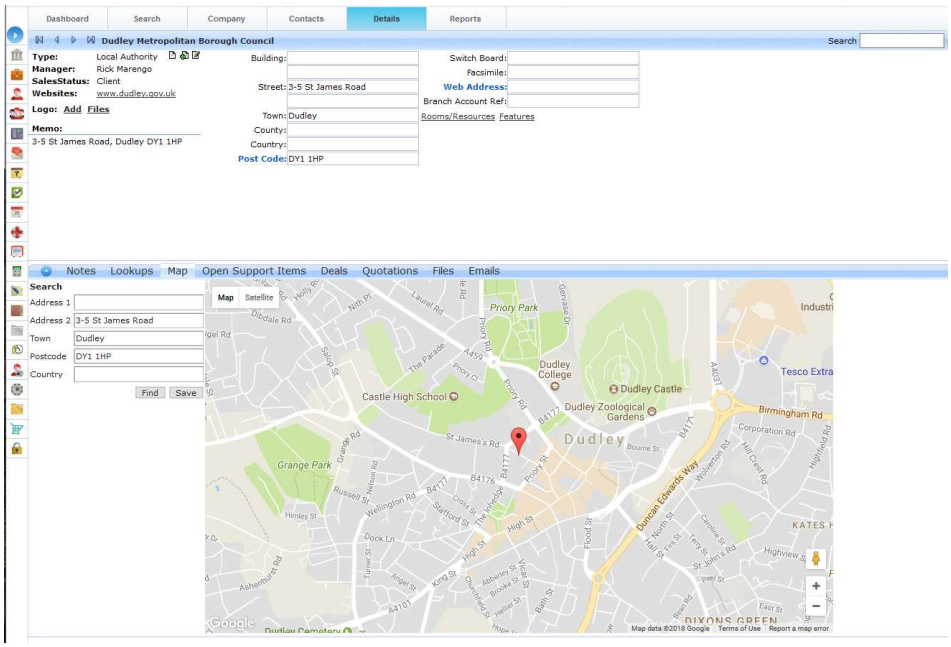
If the incoming number does not bring up the record from your CRM, then quickly copy it to the clipboard and check why that is. As soon as you search and bring up their record for editing, you can pop it in the blank space or, if it is just a number they are using temporarily, decide if it should be recorded at all or just in the notes, perhaps. For example, you might not decide to record it anywhere if it is being used because their normal device was out of order and they used an unrelated person's 'phone. On the other hand, if their phone had been handed back to a previous employer and this is the new one they will be calling with when they start to transact business from a new company, then you have a little less typing to do to add their new company, contact points, etc. The important thing is that an unrecognised number is an important point in the data quality cycle.

1.1.10 Maps

Search on maps to find your clients when visiting them. If they (or other clients you are visiting that journey) are not there then there must be a problem with their address or the Google maps Lat/ Long and that will cost your reps, service personnel and delivery drivers petrol and time



As you can see Dudley MBC is not there so I have fixed that issue in the screen below.

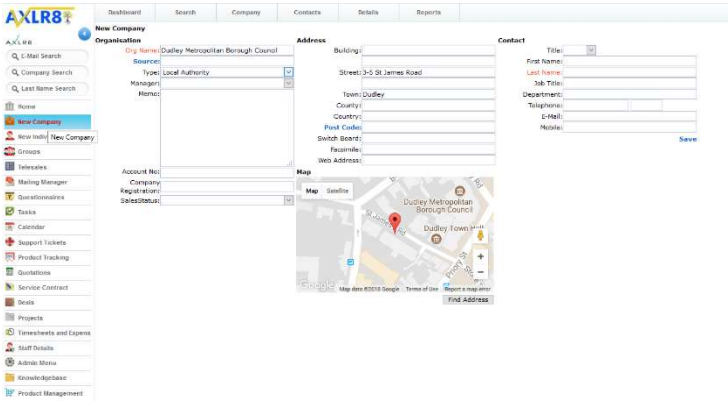


In this case it was not a data quality problem. The issue was that I typed in **St James Street, Dudley** instead of **3-5 St James Road, Dudley**.

1.2 Add companies and individuals (including profiling, etc.)

1.2.1 Type in

BY selecting the "New Company" tab at the top on the left hand menu, you can add a new organisation and the first address and contact. You can add more addresses and contacts later.



The fields should all be obvious.

1.2.2 Delegating data entry

You can provide a secure connection to a telemarketer to add more people and companies or to clean up the data you already have. Further, you can set their authority to see no data already in the system except what they have typed in. Remember, this may protect your data but they will not be

able to see if they are adding duplicates. Therefore, arrange for them to collect new data in a separate market, perhaps. If you have schools on your database, they could collect hotels.

However, if you want them to clean up data for hotels then you can provide them with visibility of only hotels in Lancashire or only hotels beginning with a-m, for example.

1.2.3 Online forms and dedupe rules

There are several options. Here ae just a couple which were mentioned above.

Clients

Customer enquiry forms generally deduplicate on emails.

Staff

Staff forms such as application forms, etc will typically deduplicate on such things as NI number, email and date of birth. For more details, please go to staffing.axlr8.com and search on AFB (application form builder).

1.2.4 Importer.

You can also import data and the way the fields are matched and deduplicated to maintain data quality is described more fully i the Importer Manual:

http://www.axlr8.co.uk/files/AXLR8DataImporterV4.pdf

Imports from disparate sources are one of the commonest causes of data quality issues if not done correctly. The above manual contains lots of common sense preparation tips and checklists.

2 Cure: Tidying up data

2.1 Data maintenance

We previously looked at adding data. This section covers updating data and deleing data.

2.2 Deletion permissions and caveats

You must have the authority to delete records. This is explained in a section below.

The deletion and all the update functions described here are audited. So, you can see the user who did them and the time stamp when they did it.

2.3 Deduplication process

The process with deduplication will need to be agreed between management and your Super User. This is our recommended flow chart.

- Decide you want to delete a record
- Make sure you are certain and double check

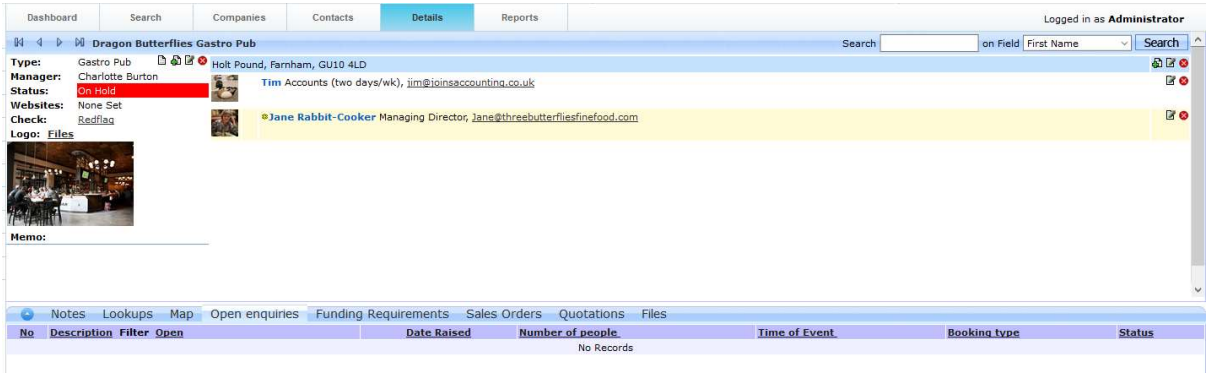
- Mark the one you want to get rid of with “XXX-“ before it in a clear way and wait for a Super User to make the deletion permanent. For example “xxx-Bloggs Haulage Limited” or “xxx-Smith, Jimmy”.
    - If a duplicate, make sure all the data is in the one you will keep before deleting the one(s) you have decided to remove.
    - In any case make sure there are no files, notes, or anything else recorded against that person/address/ contact, etc., that you will ever need again.
- Transfer, for example Support Items, Deals, etc.
    - Support items, jobs and information requests may be transferred to the person selected on the multi search screen using the link on the job.
- Delete the Contact(s) then eh Site(s) then the Organisation in that order so there are no orphaned items (referential integrity).

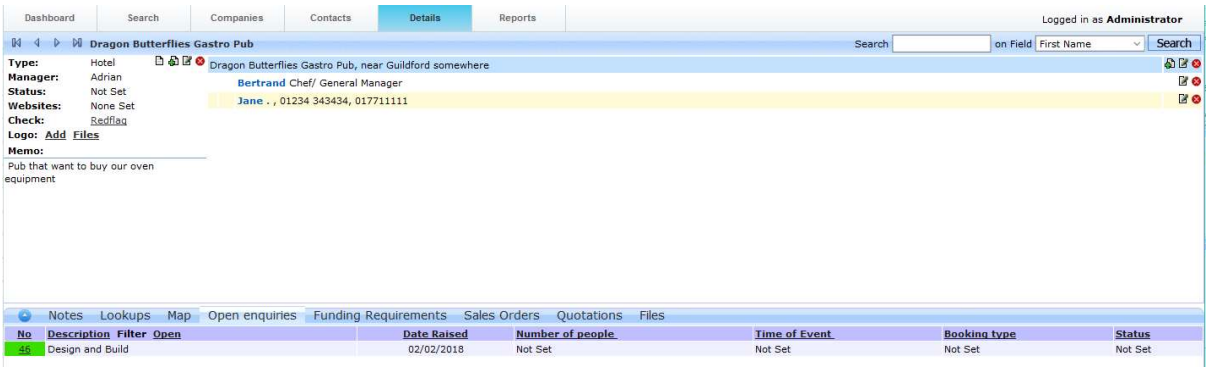
2.4 Example Process

Here is a duplicate organisation. There is also one duplicate contact (Jane).

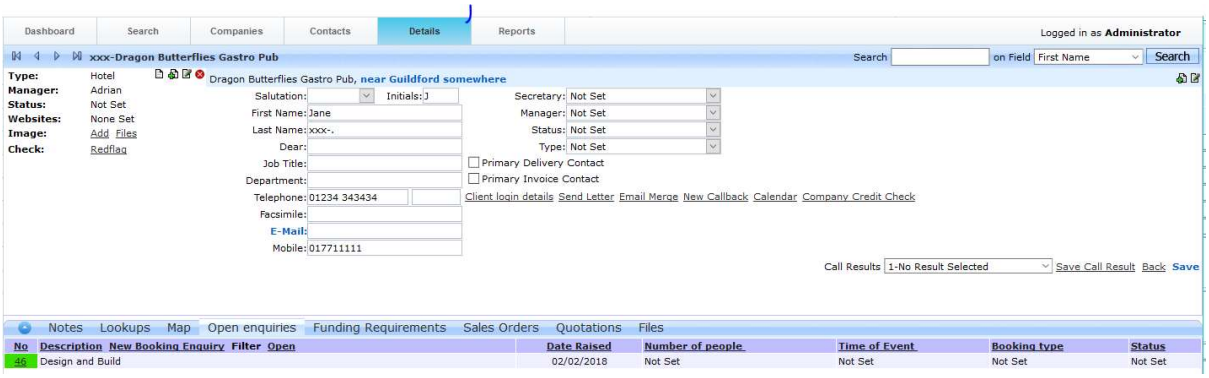
One has Jane and Jim. The other has Jane and Bertrand.

The goal is to end up with one Company record and three actual people with no duplicates.





Decide which ones are going to be deleted and mark them with "XXX-"



Before we delete anything, we need to transfer valuable information on the xxx-records (i.e. marked for deletion) to the correct records. I list those as a mobile or other telephone number we have for "xxx-jane.", any email addresses or notes, etc. I can also see an enquiry. This has ben put through on a new account by a user (Adrian) who has not checked that the account is already on there (but on hold). It maybe that Adrian is a territory Manager who is blind to Charlotte Burton's Clients including the correct Dragon Pub.

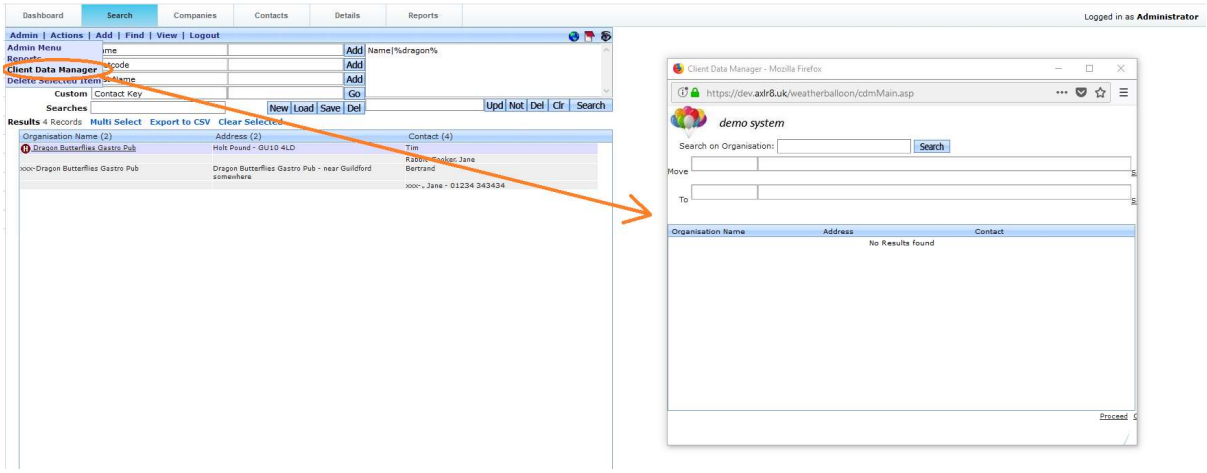
2.4.1 Transfer any jobs

2.4.2 Transferring client Contacts from one account to another

You can now transfer those contacts which are in the wrong organisation to the correct organisation.

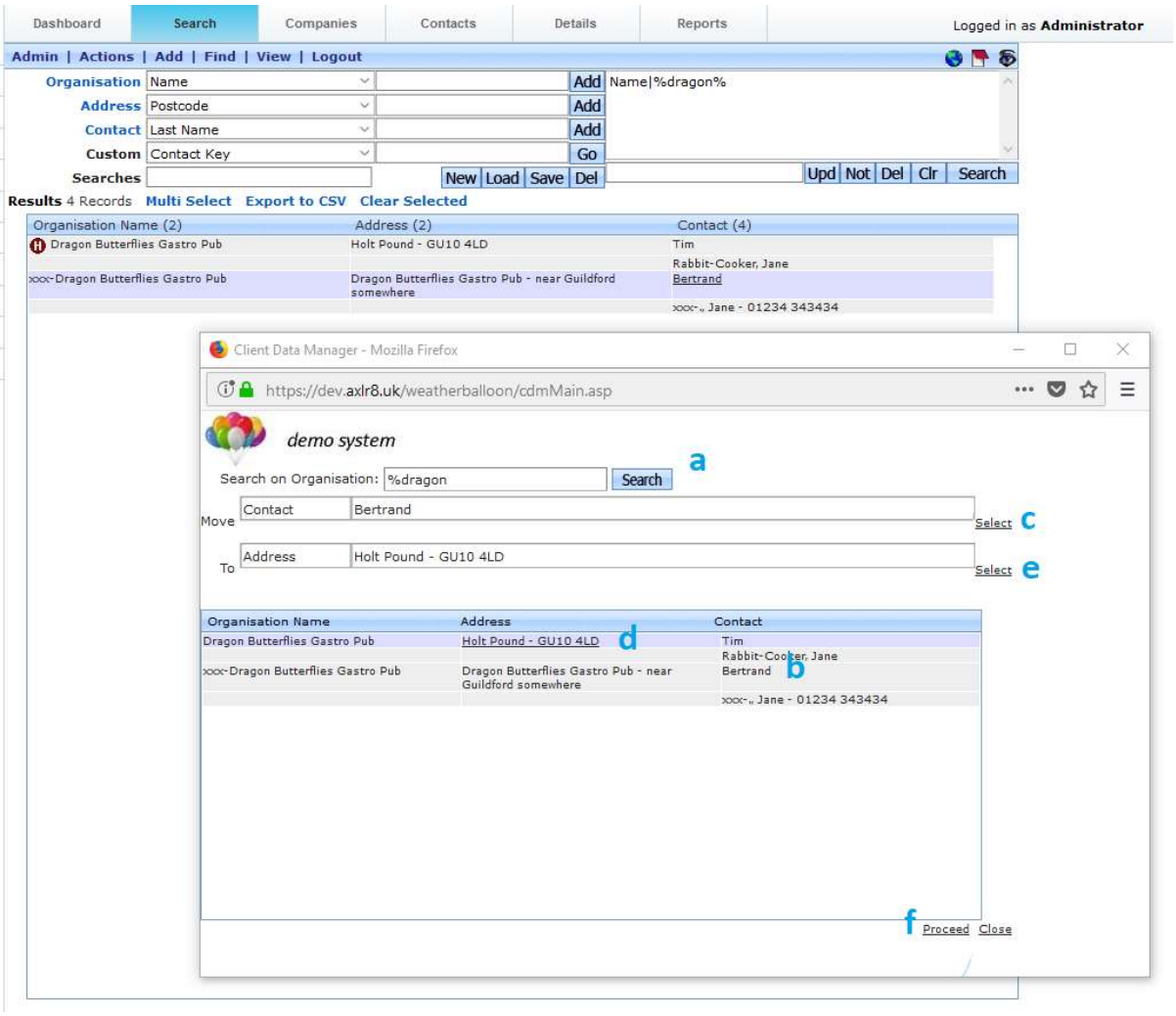
You can also transfer contacts or whole Offices between organisations. This is especially useful if organisation A with offices in Paris and London buys organisation B with offices in Tokyo, Upper Basildon and New York. Likewise, it could be two organisations merging in the public sector.

To do this, you need to have Data Management authority. This is often included in System Admin authority (see User Admin screen in the next section).



Then follow this process. Remember you must transfer a contact to an office address and not directly to an organisation. Only an office (and all its people, etc.) can be transferred to an organisation.

a. Select both or one organisations. In this case putting a wild card in front of “Dragon” will bring back the one marked for deletion as well. The wild card at the end is assumed so you do not have to type the second “%” mark although it will not do any harm.
b. Select the item to move with a single click (double click brings up all sorts of details which we do not want at this stage. In this case we chose the contact “Bertrand”
c. Click “select” and it will populate the “what to move” row.
d. Select the place you want that item to go to. In this case you can see we picked the underlined address shown.
e. Click select.
f. Lastly, a the bottom of the screen click “Proceed” and



g. deal with the warning alerts ("Are you sure?") as appropriate. It explains clearly what you would expect. check thoroughly to save time reversing the process.

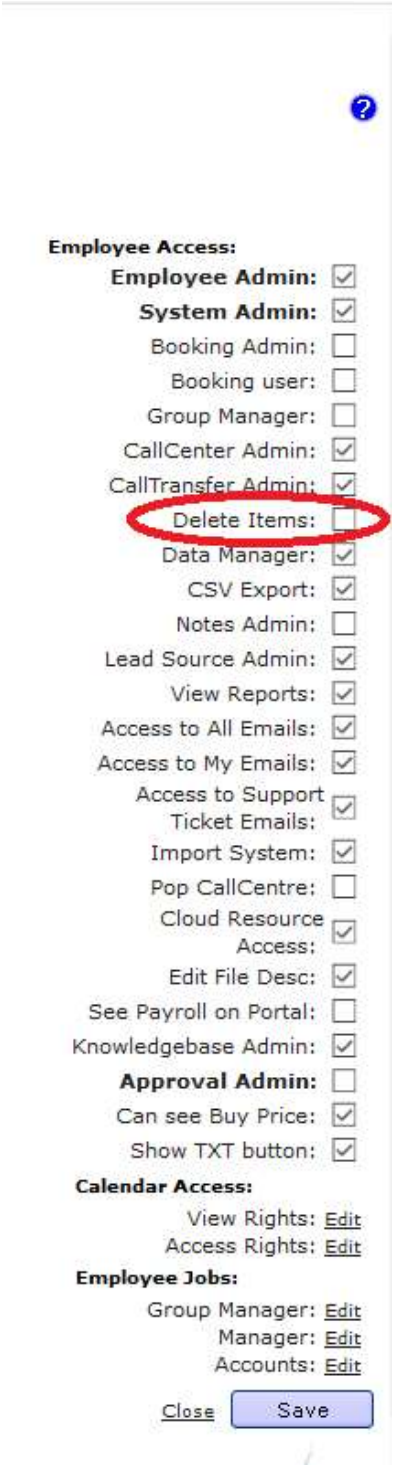
2.5 Allowing deletion for a user

The User Admin screen accessible by the Super User has an option for allowing deletion.

Delete Authority should not normally be switched on except when you are absolutely sure you want to delete an organisation, any addresses at that organisation, contact or contacts at that organisation. In the case shown above where the incoming call identified a duplicate (Sterling Co), it was obvious one needed to go. It was also obvious that no notes, support tickets, or any other items had been added to that record.

Once deletion was allowed for that person (in this case the Super User granted authority temporarily to themselves by ticking the box highlighted above), then they can delete the contact, site and organisation.

The Blue "?" question mark button provides some help documentation about all the other features seen here.

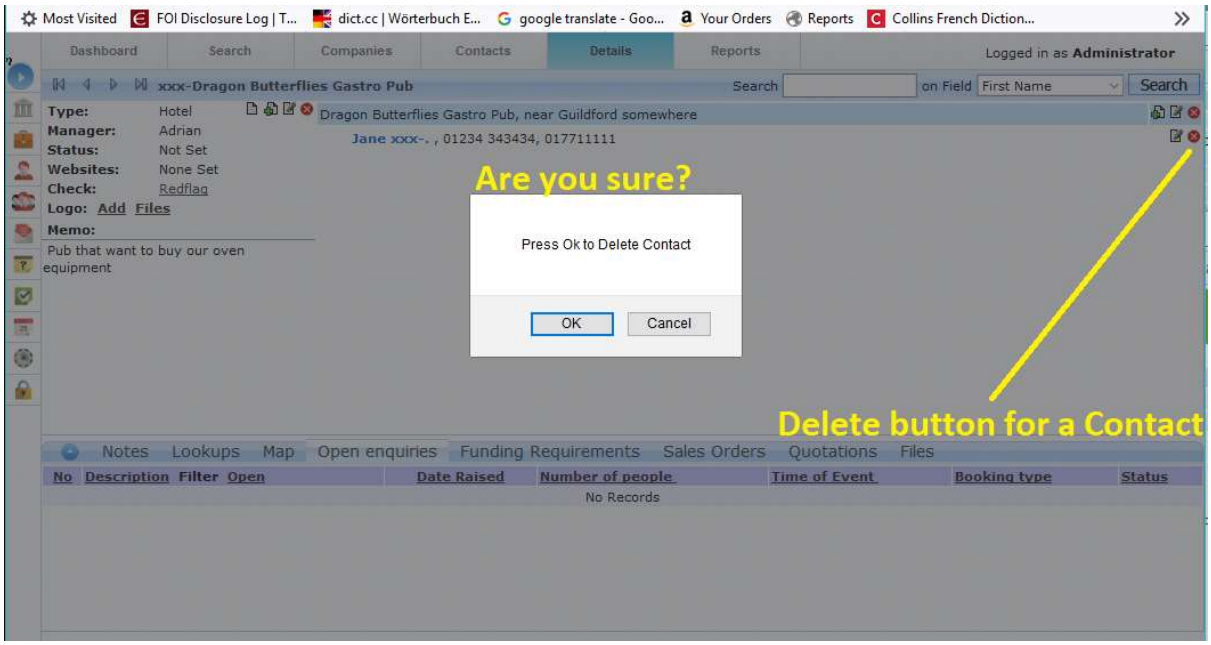


It is always important to select one of the items to select one of the duplicates (hopefully with no remaining data you do not have copied elsewhere) and mark it with XXX- in front of the name like this:

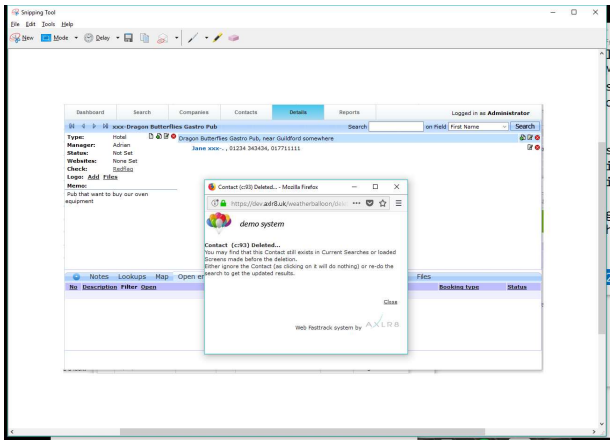
This will

a) reduce the chances of deleting the wrong one in error and
b) send it to the bottom of any searches whilst you are waiting for a Super User to process the deletion (assuming you ar not a Super User).

Then the red "Delete" command button appears for that user. First, I deleted the duplicated Contact. Then I repeated the process for the Site. Remember, there can be many contacts at a Site (branch/ office/ depot, etc.) and many sites belonging to a client organisation. Lastly, I remove the duplicate Organisation in the same way.



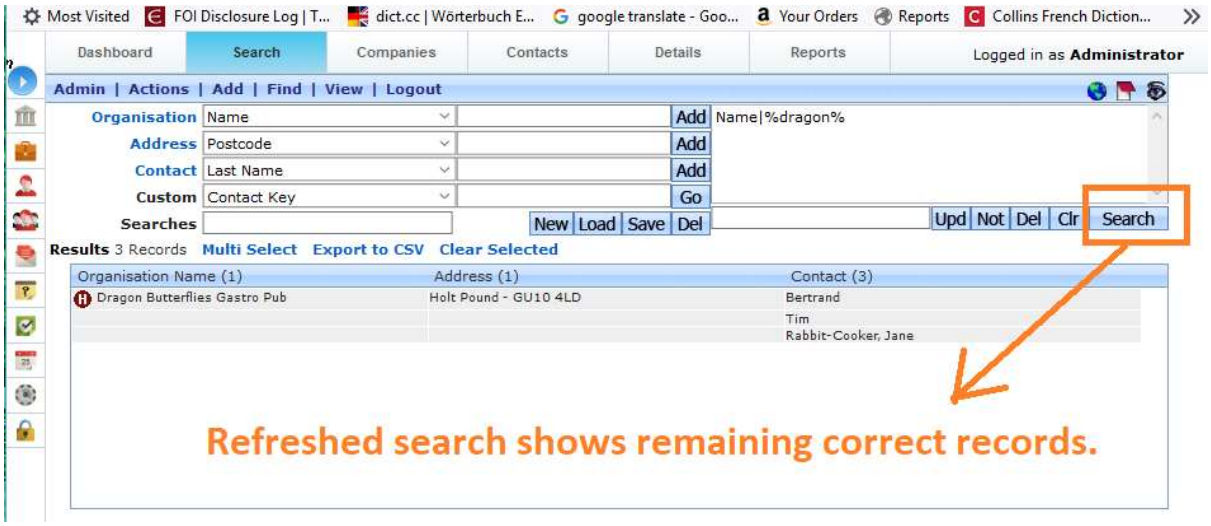
As you can see above, at every stage, there is a warning message. After every confirmed deletion, there is a confirmation screen as shown below.



Once I have completed deleting the Contact, the Site and the Organisation, I search to make sure they are no longer there. Please note that the screen you are deleting the data from will keep it there just in case you made a mistake and wish to copy and paste the data back in.

You will also have to click Search again to see that the items you have deleted have actually gone.

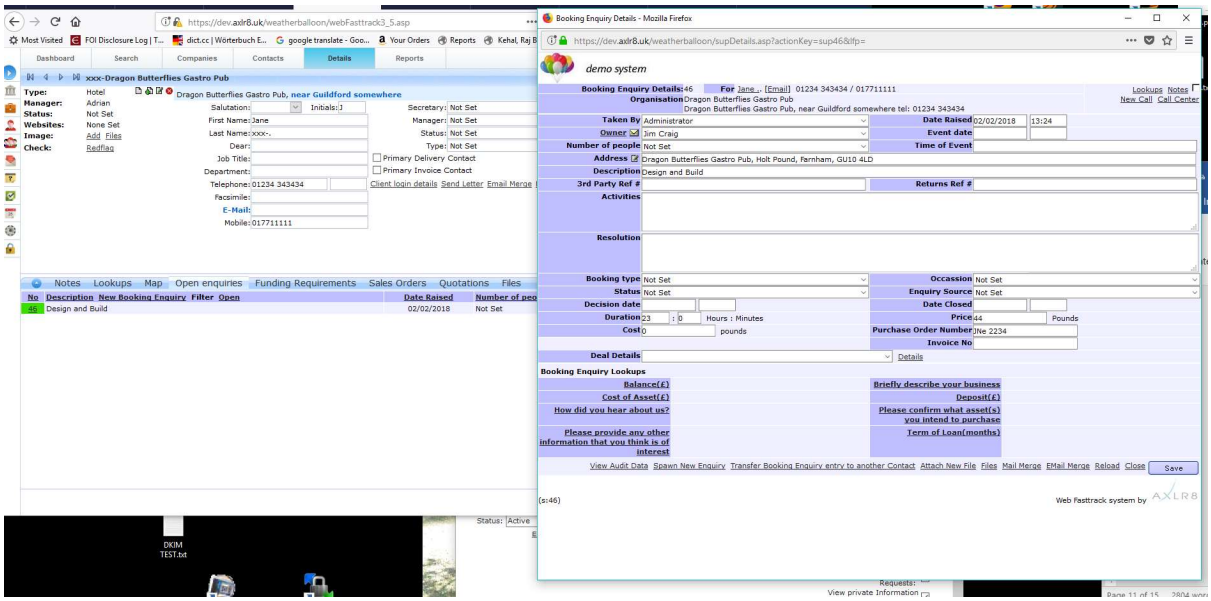
Having hit the "Search" button again to refresh, you can see the same search only brings up one organisation with the correct contacts. All their notes, any attached files (or if you use AXLR8 Exchange Vault) emails will have come across with those contacts you transferred.



2.6 Transferring jobs/tickets/requests from one contact to another

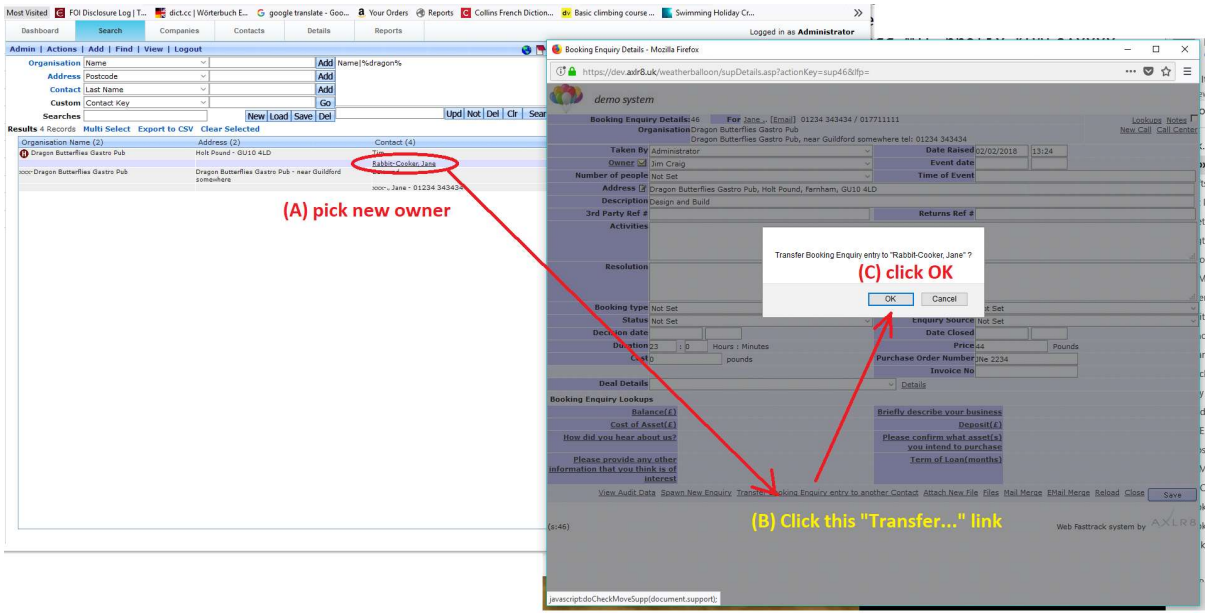
This is especially important where duplicate applicant/client contact records share the IRs between them. E.g., here we have two records for the same person called Jane. The one we want to delete has important job details on it and we want to transfer those to the record we intend to keep.

Pick the job (AKA “Enquiry”, “Support Ticket”, “Information Request” or however it is labelled for your business) and open it. As you can see at the top, it is logged to the wrong “client” contact.



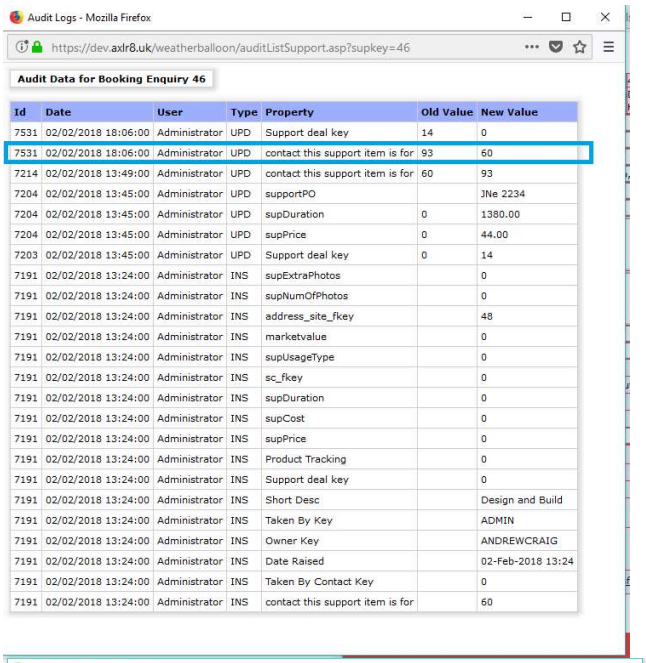
Then, as shown in the screenshot below:

- **(A)** Pick the right contact from the Search screen
- **(B)** Press the “Transfer ….” Link as shown and a popup dialogue box will come up so you can check you are transferring it to the right person.
- **(C)** Click “OK” to confirm you want to transfer that job/ticket.



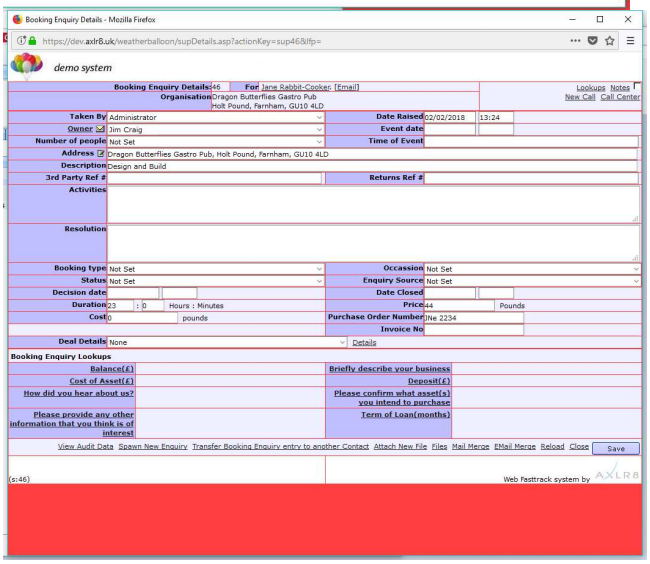
Once done, a few things may happen that you should expect.

a. The audit log will show who transferred this Job/ticket to the new Contact and when (18:06) it happened (see blue box in screenshot right). NB There is an audit entry below that where I sapped it to the wrong company to set up this demo for you at 13:49.
b. If there was a deal or service contract on the company it WAS on, then this will not come across and you need to deal with that separately. In this case a sales forecast item 14 is left on the old company. (line above the box).



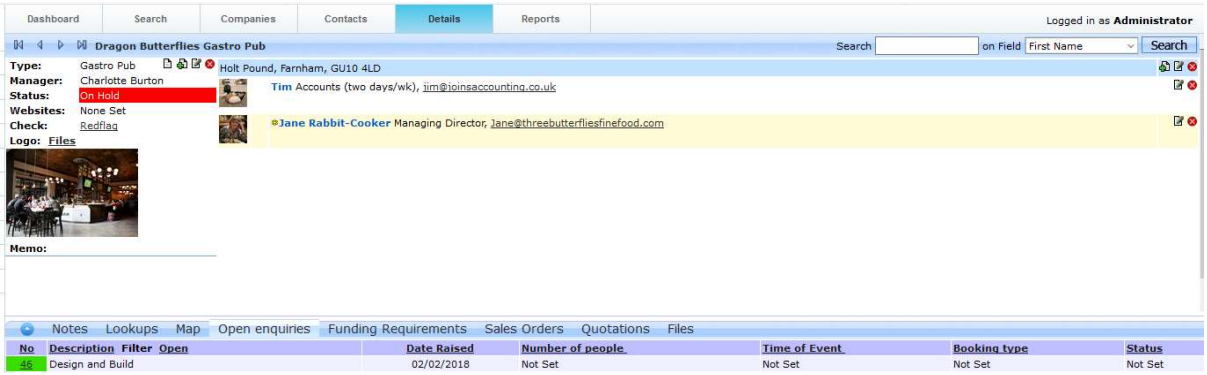
Id	Date	User	Type	Property	Old Value	New Value
7531	02/02/2018 18:06:00	Administrator	UPD	Support deal key	14	0
7531	02/02/2018 18:06:00	Administrator	UPD	contact this support item is for	93	60
7214	02/02/2018 13:49:00	Administrator	UPD	contact this support item is for	60	93
7204	02/02/2018 13:45:00	Administrator	UPD	supportPO		JNe 2234
7204	02/02/2018 13:45:00	Administrator	UPD	supDuration	0	1380.00
7204	02/02/2018 13:45:00	Administrator	UPD	supPrice	0	44.00
7203	02/02/2018 13:45:00	Administrator	UPD	Support deal key	0	14
7191	02/02/2018 13:24:00	Administrator	INS	supExtraPhotos		0
7191	02/02/2018 13:24:00	Administrator	INS	supNumOfPhotos		0
7191	02/02/2018 13:24:00	Administrator	INS	address_site_fkey		48
7191	02/02/2018 13:24:00	Administrator	INS	marketvalue		0
7191	02/02/2018 13:24:00	Administrator	INS	supUsageType		0
7191	02/02/2018 13:24:00	Administrator	INS	sc_fkey		0
7191	02/02/2018 13:24:00	Administrator	INS	supDuration		0
7191	02/02/2018 13:24:00	Administrator	INS	supCost		0
7191	02/02/2018 13:24:00	Administrator	INS	supPrice		0
7191	02/02/2018 13:24:00	Administrator	INS	Product Tracking		0
7191	02/02/2018 13:24:00	Administrator	INS	Support deal key		0
7191	02/02/2018 13:24:00	Administrator	INS	Short Desc		Design and Build
7191	02/02/2018 13:24:00	Administrator	INS	Taken By Key		ADMIN
7191	02/02/2018 13:24:00	Administrator	INS	Owner Key		ANDREWCRAIG
7191	02/02/2018 13:24:00	Administrator	INS	Date Raised		02-Feb-2018 13:24
7191	02/02/2018 13:24:00	Administrator	INS	Taken By Contact Key		0
7191	02/02/2018 13:24:00	Administrator	INS	contact this support item is for		60

c. The new (i.e. the correct) company may have different settings. So, in this case, the company is "on Hold" and that has made the job come up in red so no work is commenced. This maybe why the sales person or client created a new account! Stranger things have been known. Lucky our processes picked it up. For example, the discipline of adding a company registration should help even if the name is slightly different.



Pulls across any further information before it is lost on the duplicate contact. Now, we can delete (see below for process) "Jane xxx-." because she is a duplicate with the real Jane.

You can see the Job in the real Jane Rabbit Cooker's account "Details" as shown below.



2.7 Special cases:

2.7.1 Email bounces

If you do not deal with email bounces, you may be banned from several mail providers. To clean up mail bounces where the message clearly states that that email no longer exists, please use either the top left hand search shown above or use a string search field in a specially built report using RB.

2.7.2 Incoming calls

Incoming pop screens are a super way of checking if a client is in the database. If they are not or multiples come up, there is a data quality problem that can easily be fixed.

2.7.3 Staff database clean ups

There are various methods staffing agencies use to clean up all their thousands of contacts. Some do it manually and we can provide telemarketing and up ate facilities for this. However, without a doubt, sending staff to their portal to clean up their own data is the most productive. If you have big talent and they are too flash to do that, you will use the manual process and if you have cheap staff in very high volumes with a high turnover, then we assume you will use Newsletter Builder to remind those whose audit reports show they have not updated certain areas of data for a while that they are contractually obliged to do so. See Staff Portal manual.

2.7.4 Marketing annual opt in/ out surveys.

Again, Newsletter Builder will be a key tool for client data cleaning. Also, you can use slf service forms where they can update their details if they are club members, high volume low value consumers. If they are high value rare high net worth clients, then it's late nights and manual work if you let data quality get out of hand.

The best way to deal with clients if volumes permit is to clear things up and check as they call in or when you visit.

2.8 Other notes: third party data cleaning services

You can also hire third party data cleaning agencies to add value to your data or merge purge or clean it in other ways.

- Make sure you Export the data fromAXLR8 with our unique key/IDs
- Make sure you contract with the data cleaning agency not to lose your unique keys on the data!

Otherwise it cannot easily and accurately be returned to the database.

2.9 GDPR

2.9.1 right to be forgotten

Refer to the CSO (Contact Site Organisation) deletion above.

MAKE SURE YOU ARE NOT LOSING DATA THAT IS REQUIRED FOR SOME OTHER LEGAL OR COMMERCIAL REQUIREMENT.

For example, the tax authorities need payroll data for checking and audit purposes.

As another example, a person cannot expect to purchase an item then demand to have their record before the bill is sent out and paid by them!

2.9.2 right to correct data held

Again, easy by editing the records using the pencil icons.

2.9.3 right to see data held

This is a little time consuming but it should be a 2 minute job to send a person what you have on them in the database – including any attached files, notes, profiling data, etc. and paste it into an email