

How to Clean Your Data

For Use with
Geographic Information Systems (GIS)



ASPCA

Shelter Research & Development

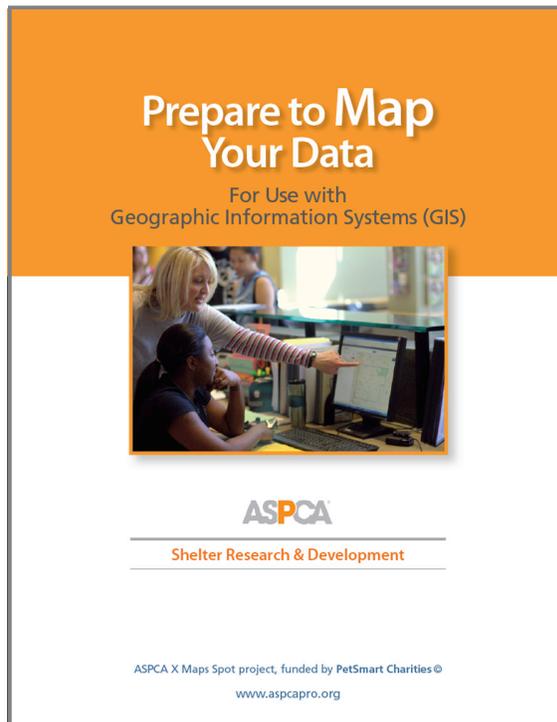
ASPCA X Maps Spot project, funded by PetSmart Charities ©

www.aspcapro.org

How to Clean Your Data

This tool was created to help determine what percentage of your data is clean enough to use for GIS purposes. Before you start, make sure you are able to view at least 85% of your community's data. This tool will lead you through the steps of cleaning intake data, but it can be used for outcome as well as spay/neuter data.

Before beginning, follow the steps in *Prepare to Map Your Data*:



Step 1: *Pull an intake report from your shelter software that includes the following information:*

- **Species**
- **Breed**
- **Intake type**
- **Intake subtype**
- **Intake date**
- **Age group**
- **Gender**
- **Altered***
- **Feral – yes or no if known** *(cats only)*
- **Found address or crossing** *(this should be the address where the animal was found)*
- **Street address** *(in PetPoint™ you want the “combined address” field)*
- **City**
- **State** *(Province)*
- **Postal Code**
- **Animal ID****



* This should be the altered status of the animal at intake, not the status when it exits the shelter. Or it should indicate whether or not the animal had surgery at your facility.

**As soon as you have your document, hide this column. You won't need it until much later, and it's easier to work with as few columns as possible.

This should be an **Excel file**, so if you can't pull a report directly to Excel you will need to **convert your report into an Excel file**.

Your document should look something like this:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	Species	Primary Breed	Intake Type	Intake Date	Age	SEX	Altered status at Intake	CROSSING	STREET_NO	STREET	STREET_TYPE	CITY	STATE	ZIP_CODE	ANIMAL_ID
1	DOG	LABRADOR RETR	OWNER SURR	01/02/2011	2 YEARS	S	INTACT		37	YACHT CLUB	DR	NORTH PALM BEACH	FL	33408	A0191368
2															
3	CAT	DOMESTIC SH	OWNER SURR	01/03/2011	4 YEARS	N	ALTERED		508	FLOTILLA	RD	NORTH PALM BEACH	FL	33408	A2374041
4	DOG	CHIHUAHUA SH	OWNER SURR	01/04/2011	2 YEARS	S	INTACT		1010	GREENPINE	BLVD	WEST PALM BEACH	FL	33409	A0191473
5	DOG	PIT BULL	OWNER SURR	01/04/2011	1 YEAR 3 M	S	INTACT		2539	WESTCHESTER	DR	RIVIERA BEACH	FL	33407	A0191491
6	DOG	CHIHUAHUA SH	STRAY	01/04/2011	1 YEAR	N	INTACT	8201 NW 24TH AVE	14352	SW 24TH STREET		MIAMI	FL	33175	A0191479
7	DOG	BOXER	STRAY	01/04/2011	3 YEARS	S	INTACT	8126 SW YACHTSMANS DR	8126	YACHTSMANS	DR	STUART	FL	34997	A0191471
8	CAT	DOMESTIC SH	OWNER SURR	01/05/2011	1 YEAR	N	INTACT		38	ALLEN	ST	WEST PALM BEACH	FL	33404	A0191515
9	CAT	DOMESTIC SH	OWNER SURR	01/05/2011	1 YEAR	S	INTACT		38	ALLEN	ST	WEST PALM BEACH	FL	33404	A0191516
10	DOG	BULLDOG	OWNER SURR	01/05/2011	1 YEAR	N	ALTERED		6772	BARRIER REEF	ST	LAKE WORTH	FL	33467	A0191507
11	CAT	SIAMESE	STRAY	01/05/2011	2 YEARS	N	INTACT	282 TIMBERWALK TRAIL	282	TIMBERWALK	TRL	JUPITER	FL	33458	A0191503
12	DOG	CHINESE SHARPEI	STRAY	01/05/2011	2 YEARS	S	INTACT	1329 LAKE VICTORIA DR	1329	LAKE VICTORIA	DR	LAKE WORTH	FL	33461	A0191504
13	DOG	LHASA APSO	OWNER SURR	01/06/2011	8 YEARS	N	ALTERED		125	INLET	WAY	WEST PALM BEACH	FL	33404	A1400228
14	DOG	MIN PINSCHER	STRAY	01/06/2011	1 YEAR	N	INTACT	100 W 20TH ST	917	SUMTER	RD	WEST PALM BEACH	FL	33415	A0191538
15	DOG	AUST TERRIER	STRAY	01/06/2011	2 YEARS	N	INTACT	1699 LAKE AVE	1020	L	ST	LAKE WORTH	FL	33460	A0191545
16	DOG	GERM SHEPHERD	STRAY	01/06/2011	2 YEARS	S	INTACT	7400 APACHE BLVD	14827	80TH	LN	LOKAHATCHEE	FL	33470	A0191536
17	DOG	RAT TERRIER	STRAY	01/06/2011	2 YEARS	N	ALTERED	4500 PALM BROOKE CIR	4899	PALM BROOKE	CIR	WEST PALM BEACH	FL	33417	A0188880
18	DOG	POODLE MIN	STRAY	01/06/2011	4 YEARS	N	ALTERED	1001 VILLAGE BLVD	5646	56TH	WAY	WEST PALM BEACH	FL	33409	A0191546
19	DOG	GREAT PYRENEES	OWNER SURR	01/07/2011	5 YEARS	S	ALTERED		110	DOLPHIN	RD	BOYNTON BEACH	FL	33435	A0191582

Step 2: Save the file frequently as you work through these steps!

There's nothing more frustrating than losing time and effort when there's an accidental system malfunction or interruption that causes you to lose data. Save often and save time.

Step 3: Consider the following six points so your data is as meaningful as possible:

1. The animal's address is what gives meaning to the GIS locator on the map.

Consider how and why your shelter staff uses different address fields. If your shelter uses the **"found address"** or **"crossing"** field to record the location where a stray animal was found, this is ideal. The names of cross streets where the animal was found are suitable for the GIS tool.

Alternatively, if your shelter is using the **"street address"** field (or fields) for the animal's address, this is the information to use. It's important that this is truly the animal's address, not the finder's address. For this reason, if records of stray animals have only the finder's address, with no indication that the animal was found at this address, these records need to be deleted from the data set.

In some shelters, it may be a combination of both – found addresses for stray animals and street addresses for all other intake reasons.

2. Animals transferred in from other shelters are not to be included in the data.

In most cases, even with local transferring shelters, there are no animal addresses. The transferring shelter's address is used. If there was an animal address, it still wouldn't be meaningful for your shelter's GIS mapping since it would be out of the area. Or if you are working in a coalition in your area, it would be a double entry of the same animal, since the same animal's information would be reported by the original agency.

3. Some intake reasons listed are not truly intakes.

Look at your intake reasons to determine which should be included in the GIS data set and which should not. For example, an intake for clinic may be an animal brought to the shelter to be spayed or neutered, or for medical treatment. A service intake may really be an animal that was microchipped only, and now the animal's information is in the shelter's system. These animals' information will also be deleted from the GIS data set.



4. Stray animals are defined as any animal picked up in the field that has no known owner.

This can include owned animals if the shelter defines owned as a stray that has been fed for a certain number of days. If you categorize an animal who has been fed for more than X number of days as owned and your GIS focus is just on free-roaming animals, you will need to include animals in the subset of "fed for X number of days" in your data set.

5. Found addresses will need to be scrubbed.

By "scrubbed" we mean that this field should show a street address or names of cross streets. Additional pieces of information in this field – such as owner in hospital, abandoned, green nylon collar, Dodge Van LP#, etc. – can't be used by the GIS mapping tool. If the additional info is after the street address it's not a problem, as the GIS software reads from the beginning of the field and stops after it has a valid street address.

NOTE: An alternative method would be to submit without a thorough scrub – the GIS software will kick back any unusable records. The disadvantage of this method is that it will drastically slow the process of analysis, since if a significant portion is unusable the set rejected by the software will need to be scrubbed and uploaded again. We can anticipate the process taking six weeks using this method.

6. When instructions include deleting rows, you should cut these rows, not actually delete them, and save them in a separate document.

By keeping track of these rows, you can get a sense of why data is being excluded and if there are ways to change your data entry methods so that you are able to delete less data.

Step 4: Sorting Round One.

With all this in mind, we can start sorting the data.

- Right click in the very top left-hand cell of the worksheet where the column and row headings meet. *This should highlight the entire worksheet and bring up a box of choices.*
- Right click on **Sort**.
- Select **Custom Sort**.
- You'll be sorting first by **intake type**.
 - » *This will group all rows of the same intake type together, making them easy to find.*
- Take a look at rows with an intake type of **euthanasia request**.
 - » *If the data includes an owner's first and last name, change the intake type to owner surrender.*
 - » *If an ACO's name or the name of a rescue group is in the owner field, change the intake type to stray.*
- Delete all rows for intake types of **clinic, service, transfer in, seized/confiscated, or other sources** that do not meet the GIS data set criteria.



The GIS tool is able to help the shelter focus on stray and owner surrender intakes. Depending on your shelter's needs, you may choose to use the tool to focus on only stray intakes or only owner surrender intakes. In this case, **delete the rows for the intake type that you are not focusing on.**

- Next, **sort by animal type or species**. Repeat the same sorting process as above.
- **Delete any rows for animals other than cats or dogs**. Your shelter may choose to use the GIS tool to focus on only one species; in that case, delete the rows for animals of the other species.
- While you have your data sorted by animal type, delete all of the entries under breed for **CATS ONLY**. Leave in the dog breeds (if you are including dogs in your data).

Step 5: Sorting Round Two.

Sort your data by address, again following the process that you used in Step 4.

- You may get this warning: *The following sort key may not sort as expected because it contains some numbers formatted as text.*

If so, select the second option – sort numbers and numbers stored as text separately and click OK.

This will move all of the records with blank address fields to the bottom of your spreadsheet. Scroll through these records.

- Delete animal records where there is **no found address or street address**. The GIS mapping tool can't use these records.
- Delete any records where there is no street address and the information in the found address/crossing field indicates that the animal was **found at the shelter**, as in the last row of the following screenshot.
- Delete any records that have a **street name but no address number or cross street** in the found address/crossing field.
- For the remaining records that have no street address, but *do* have a found address, as in the highlighted row below, you will **move the found address into the address field** in the next step.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	Species	Primary Breed	Intake Type	Intake Date	Age	SEX	Altered status at Intake	CROSSING	STREET_NO	STREET	STREET_TYPE	CITY	STATE	ZIP_CODE	ANIMAL_ID
1	DOG	LABRADOR RETR	OWNER SURR	01/02/2011	2 YEARS	S	INTACT			37 YACHT CLUB	DR	NORTH PALM BEACH	FL	33408	A0191368
2	CAT	DOMESTIC SH	OWNER SURR	01/03/2011	4 YEARS	N	ALTERED			508 FLOTILLA	RD	NORTH PALM BEACH	FL	33408	A2374041
3	DOG	CHIHUAHUA SH	OWNER SURR	01/04/2011	2 YEARS	S	INTACT			1010 GREENPINE	BLVD	WEST PALM BEACH	FL	33409	A0191473
4	DOG	PIT BULL	OWNER SURR	01/04/2011	1 YEAR 3 M	S	INTACT			2539 WESTCHESTER	DR	RIVIERA BEACH	FL	33407	A0191491
5	DOG	CHIHUAHUA SH	STRAY	01/04/2011	1 YEAR	N	INTACT	6201 NW 24TH AVE		14352 SW 24TH STREET		MIAMI	FL	33175	A0191479
6	DOG	BOXER	STRAY	01/04/2011	3 YEARS	S	INTACT	8126 SW YACHTSMANS DR		8126 YACHTSMANS	DR	STUART	FL	34997	A0191471
7	CAT	DOMESTIC SH	OWNER SURR	01/05/2011	1 YEAR	N	INTACT			38 ALLEN	ST	WEST PALM BEACH	FL	33404	A0191515
8	CAT	DOMESTIC SH	OWNER SURR	01/05/2011	1 YEAR	S	INTACT			38 ALLEN	ST	WEST PALM BEACH	FL	33404	A0191516
9	DOG	BULLDOG	OWNER SURR	01/05/2011	1 YEAR	N	ALTERED			6772 BARRIER REEF	ST	LAKE WORTH	FL	33467	A0191507
10	CAT	SIAMESE	STRAY	01/05/2011	2 YEARS	N	INTACT	282 TIMBERWALK TRAIL				JUPITER	FL	33458	A0191503
11	DOG	CHINESE SHARPEI	STRAY	01/05/2011	2 YEARS	S	INTACT	1329 LAKE VICTORIA DR		1329 LAKE VICTORIA	DR	LAKE WORTH	FL	33461	A0191504
12	DOG	LHASA APSO	OWNER SURR	01/06/2011	8 YEARS	N	ALTERED			125 INLET	WAY	WEST PALM BEACH	FL	33404	A1400228
13	DOG	MIN PINSCHER	STRAY	01/06/2011	1 YEAR	N	INTACT	100 W 20TH ST		917 SUMTER	RD	WEST PALM BEACH	FL	33415	A0191538
14	DOG	AUST TERRIER	STRAY	01/06/2011	2 YEARS	N	INTACT	1699 LAKE AVE		1020 L	ST	LAKE WORTH	FL	33460	A0191545
15	DOG	GERM SHEPHERD	STRAY	01/06/2011	2 YEARS	S	INTACT	7400 APACHE BLVD		14827 80TH	LN	LOXAHATCHEE	FL	33470	A0191536
16	DOG	RAT TERRIER	STRAY	01/06/2011	2 YEARS	N	ALTERED	4500 PALM BROOKE CIR		4899 PALM BROOKE	CIR	WEST PALM BEACH	FL	33417	A0188880
17	DOG	POODLE MIN	STRAY	01/06/2011	4 YEARS	N	ALTERED	1001 VILLAGE BLVD		5646 56TH	WAY	WEST PALM BEACH	FL	33409	A0191546
18	DOG	GREAT PYRENEES	OWNER SURR	01/07/2011	5 YEARS	S	ALTERED			110 DOLPHIN	RD	BOYNTON BEACH	FL	33435	A0191582

- While you have your data sorted by address, look at all of the records that have your shelter address or an ACO address in the address column. If there is a found address, move that to the address field; if there is no found address, these records should be deleted.

Step 6: Scrubbing.

Scroll through the rows of data:

- As you eyeball each row, look out for any blatant spelling errors in the addresses or any blanks in the city, state, and zip code fields. These will all need to be corrected.
 - » ***A quick tip to find blanks in the city, state, and zip code fields:***
Sort for each of these columns one at a time, which will drop any blanks down to the bottom of the data set, making it easy for you to see them and fill them in.
- Glance at the **intake type** and **subtype fields** – as you go through the data, you will delete records with a subtype of DOA.
- For the animals brought in as strays, look **closely at the address fields**. Several common data issues can show up here:
 - » ***There is a found address and a street address.***
Usually this shows an address for where the animal was found and the address of the ACO or individual who brought the animal in. Depending on how these fields are used by your shelter, one of the addresses should be cleared out. In most cases for stray intakes, only the found address is meaningful for GIS mapping. Thus, you will want to move the found address into the address field.
 - » ***Extra information within the found address field.***
The street number and name (or cross streets) must be at the beginning of the string. The GIS mapping tool will be scratching its virtual head if the field reads “Abandoned in Red Dodge Van, Cumberland.” However, the mapping tool will ignore any information that appears after the address. So, “Found at Oak St and Holly Ave” will not work, and the “Found at” will need to be deleted, whereas “Oak St and Holly Ave – found yesterday” is acceptable.
 - » ***There is an address listed, but there is no information in the found address/crossing field indicating that the address is where the animal was found.***
In this case, the address listed is likely the finder’s address, not the animal’s address, which is not useful for this purpose. These records should be deleted.
 - » ***There is a landmark listed instead of an address or intersection in the found address/crossing field.***
If you would like to keep these records in your data set (which you likely will, if there are a lot of them), you will need to use Google Maps to look up the address of the landmark and then fill in that information. So, in the case of the highlighted record below, you would look up Best Buy, and fill in the address, city, state, and zip code in the appropriate fields.

See the screen capture on the next page.

	C	D	E	F	G	H	I	J	K	
	Intake Type	Intake Date	Age	SEX	Altered status at Intake	CROSSING	STREET_NO	STREET	STREET_TYPE	CITY
1	OWNER SURR	01/02/2011	2 YEARS	S	INTACT		37	YACHT CLUB	DR	NORTH F
2										
3	Stray	01/03/2011	4 YEARS	N	ALTERED	Found at Best Buy on 1st St				NORTH F
4	OWNER SURR	01/04/2011	2 YEARS	S	INTACT		1010	GREENPINE	BLVD	WEST PA
5	OWNER SURR	01/04/2011	1 YEAR 3 M	S	INTACT		2539	WESTCHESTER	DR	RIVIERA
6	STRAY	01/04/2011	1 YEAR	N	INTACT	6201 NW 24TH AVE	14352	SW 24TH STREET		MIAMI
7	STRAY	01/04/2011	3 YEARS	S	INTACT	8126 SW YACHTSMANS DR	8126	YACHTSMANS	DR	STUART
8	OWNER SURR	01/05/2011	1 YEAR	N	INTACT		38	ALLEN	ST	WEST PA
9	OWNER SURR	01/05/2011	1 YEAR	S	INTACT		38	ALLEN	ST	WEST PA
10	OWNER SURR	01/05/2011	1 YEAR	N	ALTERED		6772	BARRIER REEF	ST	LAKE WC
11	STRAY	01/05/2011	2 YEARS	N	INTACT	282 TIMBERWALK TRAIL				JUPITER
12	STRAY	01/05/2011	2 YEARS	S	INTACT	1329 LAKE VICTORIA DR	1329	LAKE VICTORIA	DR	LAKE WC
13	OWNER SURR	01/06/2011	8 YEARS	N	ALTERED		125	INLET	WAY	WEST PA
14	STRAY	01/06/2011	1 YEAR	N	INTACT	100 W 20TH ST	917	SUMTER	RD	WEST PA
15	STRAY	01/06/2011	2 YEARS	N	INTACT	1699 LAKE AVE	1020	L	ST	LAKE WC
16	STRAY	01/06/2011	2 YEARS	S	INTACT	7400 APACHE BLVD	14827	80TH	LN	LOXAHAT
17	STRAY	01/06/2011	2 YEARS	N	ALTERED	4500 PALM BROOKE CIR	4899	PALM BROOKE	CIR	WEST PA
18	STRAY	01/06/2011	4 YEARS	N	ALTERED	1001 VILLAGE BLVD	5646	56TH	WAY	WEST PA

Step 7: Sorting Round Three.

Now we can clean up the age column, if necessary.

The GIS mapping tool will be looking for “juvenile” or “adult” in this field. Many shelters use age brackets, kitten, or puppy instead of juvenile or adult. There may be other variations.

- Right click in the very top left-hand cell of the worksheet where the column and row headings meet. That should highlight the entire worksheet and bring up a box of choices.
- Right click on Sort.
- Select Custom Sort.

This time you’ll be sorting by age group.

- The age separation for juveniles vs. adult is 5 months. For example:
 - » replace all “kitten” or “0 – 5 mos.” with the word juvenile.
 - » Replace ages that are higher than 5 months (i.e. 2 years) with the word adult.
 - » Replace all age brackets or words for adults such as “dog,” “cat,” or “senior,” with the word adult.
 - If the word unknown is in this column, leave it;
 - if the field is blank, fill it in with the word unknown.

NOTE: Some shelters use 6 months as a cut-off for adults vs. juveniles and are unable to separate out only animals that are 5 months or younger.

Step 8: Double-Checking.



Now that your data is in place, let's do a spot-check.

- Unhide the column with animal ID #s.
- Choose a few rows at random.
- Enter the ID #s from the chosen rows back into your shelter software to make sure that the information in your Excel spreadsheet matches up to the information in your shelter software.



For more information:

www.aspcapro.org/gis