COMMON GROUND™ BEST PRACTICES GUIDE

A complimentary resource exclusively for Convio Common Ground clients

Data Hygiene and Duplicate Management Best Practices

Written by Heller Consulting

July 2010

HELLER CONSULTING
your fundraising technology team

convio
MOVE PEOPLE™

Contents

About the Author

This guide was written by Heller Consulting, a Convio Solution Provider authorized to provide services around the Common Ground system. It is part of an ongoing effort by Heller Consulting and Convio to get information about nonprofit technology and fundraising and marketing best practices into the nonprofit sector's hands. Heller Consulting brings efficient business practices and effective use of technology to nonprofit fundraisers, supporting greater success with less effort. Since 1996, the company has served more than 700 nonprofits of every type and size. For more information, please visit www.teamheller.com.

I. Introduction

"Dirty data" sounds bad. It sounds like something an organization would want to avoid. Yet, dirty data is a nearly universal problem for nonprofits. Since 1996, we at Heller Consulting have encountered few organizations whose fundraising databases had no problems stemming from dirty data at the onset of our work with them.

What does "dirty data" mean? For our purposes in this guide, it means duplicate records and records with little to no useful information that are just taking up space. The consequences of dirty data are serious: Higher costs (in printing, postage and staff time, for example) to use and maintain bad data, which translates into lower income for your mission. Dirty data also has repercussions for donor relations – it's a rare fundraiser indeed who has never sent two pieces of mail to the same address or scoured a mailing list for duplicates before using it.

Dirty data has numerous causes and no easy solutions. An organization's best protections against dirty data are focus on the issue, vigilance, clear procedures, and thorough knowledge of the tools available to remedy the situation. Of these, the greatest is making data hygiene a priority and recognizing data hygiene as a necessary, valuable part of system administration.

This guide is intended for system administrators at organizations of any size who are either struggling with duplicate management or are wondering if their system has a data quality problem. There are three sections in this guide: how to bring clean data into your Convio Common Ground system, how to merge duplicates that are already in the system, and how you can assess the quality of your data.

Our goal is for you to be able to determine the scope of your data issues and identify the best tool to resolve the issues. Other references exist for training on the tools we mention. Throughout this guide, we'll reference resources for training or more information about the tools we are recommending.

After reading this guide, you should be able to:

1. Describe the most popular import tools and know which tool to use for an import.
2. Know how to apply the available duplicate management tools.
3. Create a table of contents for an internal "data hygiene" guide.
4. Navigate and describe the tools for managing Common Ground-Convio Online Marketing data in a synchronized environment.
5. Determine the scope of your organization's duplicate problem.

Please note that our Acknowledgement and Stewardship Best Practices Guide discusses in greater detail the importance of the quality of information on records, such

as addressees and salutations and verified addresses. That guide is available on the Resource Center tab in Common Ground.

II. Getting Clean Data into Common Ground Cleanly – Did We Mention *Clean*?

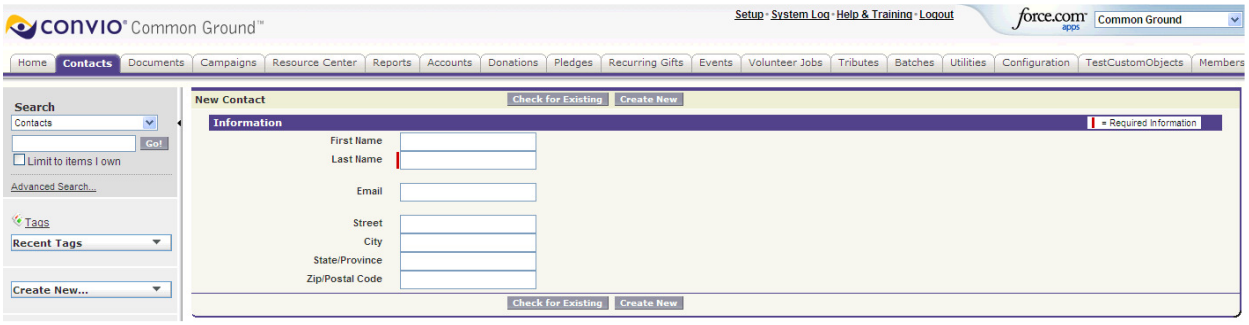
There are four ways that records come into your Common Ground system: manual entry, import, the Convio Common Ground Connector, or some other automated integration between Common Ground and another system. In this section, we'll address best practices and tips for bringing clean data into the system in the first place.

A. Manual Entry

For some organizations, manually keying new records into Common Ground is the primary way new records get into the system. For other organizations, it may be the exception if imports and/or the Connector are in heavy use. In any case, it will always be an option, and it's important to set best practices and know how to use all the native functionality Common Ground has to offer for ensuring clean data upon entry.

One of the most important safeguards against adding duplicate records to the system is to always search first to see if a Contact exists before adding it. Make sure your users know how to use the most effective method of searching for Contacts on the Contacts tab.

To get to this screen: On the Contacts tab, click on the **New** button, and then click the **Continue** button as if adding a new Household Contact.

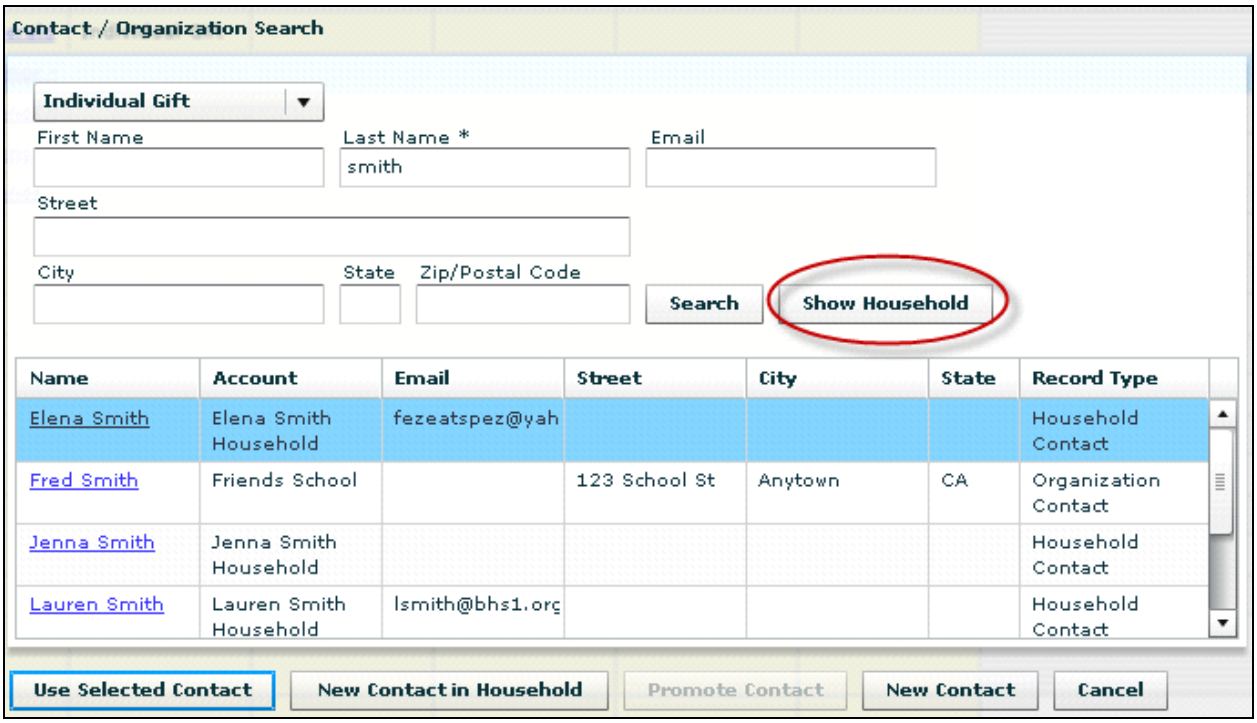


Enter the information about the potential new Contact, and then click on **Check for Existing**. Note: Even though on the previous screen you said you were adding a Household Contact, Common Ground also checks for existing Organization Contacts via this search method.

Also, it's usually preferable to add a Contact to an existing Household where applicable (such as in the case of a spouse or other household member) than to create a new Household.

When completing a new contact record, it's worth the effort to fill in as much information as possible. This can be especially useful in the future should it become necessary to evaluate a record as a potential duplicate, or for any projects such as a wealth screening. For example, patronymic suffixes (Jr., Sr., III) can make it clear if two records with the same name are the same person or not. It's also worthwhile to make sure relationships or record origin codes are present, to give an indication (in the absence of a gift history) of why the record was added to your system. That could make a difference in the future if there is a need to assess the value of records in the system.

When using Batch for gift entry, always search first to see if a Contact exists before adding a new donor. In addition, you can take the additional step of double-checking Households to see if the donor should be added to an existing Household. This is done through the use of the "Show Household" button on the Contact search screen when looking up a donor in a gift batch. Clicking this button narrows the search screen to contacts associated with the Household in the highlighted row.



After double-checking the Household, the user is in a better position to decide if the donor should be added to an existing Household or if a new Contact with their own household should be created.

B. Importing Records

There are several tools available for importing records into Common Ground. Importing is a popular option for adding records to the system, as it is much faster than hand-keying data, and it is easier to massage large quantities of data into shape for best results. In this section, we'll discuss four of the most prevalent import tools, and when you should use each one.

Before you even think about import tools, however, be sure that the source data file is in the best shape it can be. Typically, source data files are in comma-separated (.csv) or text (.txt) format. Each column in the source data corresponds to a field in Common Ground. For example, if the source data file has one column with the person's name in it (John C. Doe), this will need to be broken down so that there is a separate column for First Name, Middle Name and Last Name. Title and Suffix would also be separate columns. Convio offers data import templates (http://customer.convio.com/cgdataimport) that demonstrate the correct layout for many common types of imports.

Next, make sure there are no duplicates in your file, or between multiple files that will be imported at the same time. If your file can be opened in Excel, use sorting, filtering and subtotals to check for duplicates. In Excel 2007, there is a Remove Duplicates function in the Data menu.

You can also write a simple formula to determine if two consecutive rows are identical. In the example below, a new column was inserted in Column A for the formula and then copied down the length of the file.

A2 | fx | =IF(B3&C3&D3&E3&F3&G3&H3=B2&C2&D2&E2&F2&G2&H2,"Dupe","")

	A	B	C	D	E	F	G	H
1	Dupe?	First Name	MidName	Last Name	Address	City	State	ZIP
2	Dupe	John	K.	Palmer	4990 Doe Meadow Drive	Annapolis Junction	MD	20701
3		John	K.	Palmer	4990 Doe Meadow Drive	Annapolis Junction	MD	20701
4	Dupe	Denise	D.	Frost	721 Godfrey Street	Burlington	OR	97203
5		Denise	D.	Frost	721 Godfrey Street	Burlington	OR	97203
6	Dupe	Mary	R.	Christian	4826 Brighton Circle Road	Correll	MN	56227
7		Mary	R.	Christian	4826 Brighton Circle Road	Correll	MN	56227

Then, perform a Copy → Paste Special on column A to paste only the values of the cells, and get rid of the formula. Sort by column A and delete the rows that say "Dupe." Then, delete column A.

When comparing records between multiple files, the VLOOKUP formula in Excel is helpful, especially if the Contact ID is present in both files. With VLOOKUP, one file can be checked against another file for matches, provided the Contact ID is present in

column A of both files. Files can also be brought into Access and compared using a Select query. When using any of these methods, it's important to compare apples to apples. That is, the columns in each file should de defined the same way (although the column headers do not have to match).

Identifying Existing Records

It's generally important to consider whether a given import file could have existing Common Ground contacts in it. Failure to be aware of existing contacts in an import file is a leading cause of accidental import of duplicate records. Ideally, your file will have a column for Contact ID, which clearly separates existing records from potential new records.

Experienced database administrators always include the Contact ID when exporting records from the system, in case an update (especially an unforeseen update) is planned or becomes necessary. For example, it is not uncommon for a mailing list to be exported from Common Ground for review before the list is sent to the mail house. By including the Contact ID, it becomes possible to add the Campaign Membership only to those records that remain in the final draft of the mailing list.

If your file does not have an easy way to tell potential new records from known existing records, *you will need to choose an import tool that will do this for you* (such as PeopleImport, described in the next section). It is **not** necessary to look up each record by hand.

One common question is whether new records (imports or inserts) need to be processed separately from existing records (updates), if the two types of records are in the same file. The answer depends on which tool you are using, and whether or not a record identifier (such as the Contact ID) is present. Some tools can do "upserts", which process updates and imports in the same file, and some cannot.

The Right Tool for the Job

In this section we will discuss four top import tools that are in widespread use by Common Ground clients: DemandTools™, PeopleImport™, Apex Data Loader™, and Salesforce.com's own Mass Import Contacts feature.

The following table summarizes the most critical and relevant features for Common Ground administrators that are available with each tool.

	Does upserts	Set customized duplicates criteria	Uses System IDs for dupe checking	All Contact fields available	All Objects available	Saves import templates
DemandTools				X	X	X
PeopleImport	X	X	X	X		X
Mass Import Contacts	X		X			
Apex Data Loader	X		X	X	X	X
Manual entry	X	X		X	X	

This grid can help you decide which import tool will best meet your needs. Ask yourself the following questions. Your answers will help narrow the list of tools to the best one for the job.

- Is the same import going to happen again (for example, a regular upload from a lockbox file)? If so, look for a tool that will save import templates.
- Does the file contain IDs that will identify existing records? If not, look for a tool that can use alternate fields to identify existing records.
- Do you want existing records to be updated, or treated as exceptions? If existing records should be updated, look for a tool that will do upserts.
- Which objects will be affected by your import? Some tools only import to Leads and Contacts, while others can import into all objects.
- Are there lookup fields in your file (such as Classification Type)? If so, look for a tool that can handle all Contact fields.

Now let's discuss each tool in detail, and consider some situational examples that best fit each tool. You may notice that no tool has all the features that are mentioned in the grid. Where possible, we note workarounds for when the tool will not meet all of your needs.

1. DemandTools and PeopleImport

CRMfusion, Inc. is a Canadian software company that specializes in data management applications for Salesforce databases (www.crmfusion.com). They offer three powerful and complementary data hygiene-related applications: DemandTools, PeopleImport and DupeBlocker™. All three products are free to nonprofit organizations. DemandTools comes with one copy of PeopleImport. We'll discuss DupeBlocker in the next section on duplicate management.

PeopleImport is probably the most sophisticated import tool available to Common Ground administrators. PeopleImport can be used only to import Leads and Contacts —

it cannot import to any other objects or be used to update existing records. PeopleImport walks the user through a five-step wizard in order to import records. One of these steps is to identify duplicate criteria, which PeopleImport can use to stop duplicate records from being imported. Due to these customizable duplicate criteria, it's not required to have a record identifier in the file (although it is preferable).

DemandTools is a powerful application that is popular with advanced database administrators. There are literally hundreds of ways DemandTools can be used to maintain a Common Ground database, including duplicate resolution, which will be discussed in the next section.

DemandTools is a robust tool that requires training and experience to use properly. However, when used properly, organizations get a lot of value from it.

When importing records, DemandTools is most effective if you are certain that there are no existing records in your file. This is because the "MassEffect" module of DemandTools will import data, but it will not check to see if records already exist. That step is performed by the "Find/Report IDs" module, which will produce a report of existing records in the file, but does not change data. So, it would be a two-step process to identify existing records and then remove them before importing the new records. The MassEffect module can import data to objects other than Contacts and Leads. If there is a chance that existing records are in your file, and you only need to work with Contacts or Leads, PeopleImport is a better choice.

Many administrators describe DemandTools as "indispensable." But, beware: It's easy to make major changes to your data, which can be disastrous if the application is not used correctly. CRMfusion has video tutorials on their website, and once you have the license for DemandTools and PeopleImport, you can access several free trainings offered each month (www.crmfusion.com/demandtools/details.php). We also recommend that administrators new to DemandTools practice in a test environment until they are comfortable performing mass changes and imports.

2. Apex Data Loader

Apex Data Loader is located within Common Ground/Salesforce in Setup → Administration Setup → Data Management → Data Loader. To access and use it, you must have a profile with "Modify All Data" permission. It's probably the most straightforward tool on this list for doing straight imports from external files. There is not as steep a learning curve as for DemandTools, making it a popular choice for administrators who prefer more user-friendly tools. Also, the Data Loader can access other objects besides Leads and Contacts.

The Data Loader identifies existing records *only* by a Contact ID or a custom field set up as an External ID. If you use it to import records from a file that does not have a column for the record identifier, it can be easy to create duplicates. In that situation, it would be better to use PeopleImport. The Data Loader can perform an upsert if the file has a column for a record identifier and is a mix of new and existing records. An "upsert" will insert new records when the record does not already exist, and will overwrite information on existing records when a match is found.

3. Mass Import Contacts

Mass Import Contacts is the native Salesforce import tool available inside Common Ground/Salesforce. It is located in Setup → Data Management → Import Accounts/Contacts. Mass Import Contacts has an easy, intuitive interface for loading and mapping the file (must be a .csv file), but there are some limitations. It is best used as an import/update tool for files that include Contact ID, Account ID, or Account Name for existing records. These are the **only** fields Mass Import Contacts recognizes as record identifiers. It does not check anything else to identify a potential duplicate. Another limitation is that lookup fields (such as Classification Type) cannot be mapped, with the exception of Record Owner. In addition, this tool does not allow you to save "templates" of commonly used imports for when you may have the same file format coming to you on a regular basis.

Convio's Data Migration Templates

Convio provides many free data migration templates that your organization can use to design your own import files (http://customer.convio.com/cgdataimport). These templates are designed to work with Apex Data Loader. If the column headers on these templates are left intact, they should map automatically to the correct field. The templates are distributed in .xls format so that notes and tips about each column/field can be included, but must be saved in .csv format before being used for import. Required fields are indicated. The following templates are available:

- Accounts and Contacts
  - Organization Account
  - Organization Contact
  - Household Account
  - Head of Household Contact
  - Household Contact
- Donations
  - Single Donations
  - Pledges
  - Pledge Installments
  - Recurring Gifts
  - Recurring Gift Payments

    - Donation Designations
- Relationships
- Event Management
    - Events
    - Levels
    - Invitations
    - Attendees
    - Teams
- Volunteer Management
    - Jobs
    - Shifts

Also included with the templates is important information about the recommended order in which to utilize the various templates for performing multi-object imports (see situational example #1 below).

Situational Examples

There are some typical situations that administrators encounter when importing data into Common Ground. This section provides a brief overview of how to handle them.

1. **Contacts plus another object in the same file.**
   A file with Contacts plus some other piece of information (Donations, Activities) is probably the most prevalent type of multi-object import. This is an important situation to be aware of, since data can only be imported into Common Ground (or any SalesForce database) one object at a time.
   a. Import the Contacts first with PeopleImport. (Recall that Convio offers templates to help design the file if needed.) There is no need to manipulate the data file — simply leave the other columns out of the mapping and they will be ignored.
   b. PeopleImport produces a “success” file that includes all the columns of the original file plus the IDs of the imported records.
   c. Import the other object’s information with Data Loader or DemandTools. Again, simply do not map unnecessary columns and they will be ignored. Either tool will work with Convio data migration templates if you need help finalizing your file layout.

2. **A custom object is involved.**
   We wanted to call out this situation because there will not be any Convio data migration templates for the custom object. You will need to figure out the fields to have in your file. However, the column headers do not need to be exact. You can manually map the column headers to the fields. If the file is Contacts plus information for a custom object, it will be similar to situation #1.

3. **Your boss gives you a file of names and addresses to make sure they are in the system. How do you add the new names without importing any duplicates?**
    - PeopleImport will have the broadest capability to check the system for existing records. This capability depends on the quality of the duplicate criteria that you set up.
    - Before beginning, you can choose to:
        - Insert new records only (existing records will be kicked out as exceptions), or
        - Update existing records only (new records will be kicked out as exceptions), or
        - Both insert and update. Generally, it would be considered more conservative to allow only new records in, and then process the updates separately, after having time to examine the records to make sure they are updates.

4. **Add a Campaign Membership to records that were in the final version of a mailing list.**
   This situation arises when a mailing list is exported from the system without using the Segmentation features. Typically, this is a general file that is passed around the office and reviewed by different stakeholders, who can choose to remove names from the list. The final file is then used to produce the mailing, and a record of each person who received the mailing must be put into the database.
    a. Make sure your file has the Contact ID when you export it from the system.
    b. Use DemandTools or DataLoader to upload the Campaign Membership for only those records that are in the final version of the file.

C. Convio Common Ground Connector

Convio Online Marketing Duplicate Management

When Common Ground and Convio Online Marketing (COM) are synced via the Connector, it is especially important to manage the quality of the data in both systems. Starting with your implementation, make sure you and your implementation partner are following best practices for setting up the Connector. Two best practices are especially important:

- All duplicates should be cleaned up in both systems (COM and Common Ground) before turning the Connector on. However, we recognize that this may not be a realistic standard for many organizations. At a minimum, for Common Ground implementations that include a new instance of COM, there should not be Contacts in Common Ground with duplicate emails when the initial

synchronization is performed. If there are records with duplicate email addresses in Common Ground, the situation will be perpetuated into COM, and multiple copies of the email will be sent to the same email address in email blasts where both records are included.

- Ideally, if COM was already being used before the Common Ground implementation, there should be no overlap between the data in both systems. That is, every record in COM that can be matched to the corresponding record in Common Ground should be matched before the Connector is turned on. Otherwise, duplicate records will be created in both systems. Convio provides a service called "equalization" to reconcile COM and Common Ground before the Connector is turned on for precisely this purpose. This process requires your organization's involvement and must be completed before the Connector can be activated.

On an ongoing basis, be sure to keep duplicates within each system to a minimum so that the duplicates are not inadvertently perpetuated into the other system.

COM has functionality called "Resolve New Registrations" that will hold new contacts identified as potential duplicates in quarantine. These potential duplicates must be addressed before these records (and their associated transactions) can be part of the Connector sync. However, the reverse is not true for data flowing from Common Ground into COM. That is, there is no quarantine to stop new duplicate Common Ground records from being loaded into COM during a Connector sync.

Check the Resolve New Registrations area regularly to keep up with potential duplicates; the "Resolve New Registrations" area can get backlogged if this maintenance is not performed regularly. If a backlog happens, you can address it in two passes:

1. On the first pass, look for pairs of records where either one or neither record has a "Member ID" (Common Ground ID). Merge these records or mark them "Not A Match."
2. On the second pass, look for pairs where both records have a Member ID and both records are still in Common Ground. Resolve these duplicates in Common Ground. The merges will be carried into COM during the next Connector sync (see below for more information about this functionality).

Recent improvements to COM duplicate detection and resolution reduce the number of pairs that need to be reviewed by a user, so the process will take less time and be less prone to user error. We summarize them here, and more information is available in the Connector guide (http://community.customer.convio.com/docs/DOC-2027) and in the Common Ground community (http://community.customer.convio.com/docs/DOC-2745).

- There are fewer false positives, defined as obviously different people who share the same address. The system is smarter about recognizing new people and will automatically allow new records to be created.

- There are fewer obvious positives, meaning the system is smarter about recognizing matches and will not show as many obvious matches.
- A new feature checks to see if two individuals are members of the same household before flagging them as potential duplicates.
- You can set a configurable threshold, above which records between COM and an offline database can automatically be matched.

Common Ground: The Buck Stops Here

One significant change in Common Ground/COM integrated duplicate resolution is that in version 2.5 and above, merges of records that are in both Common Ground and COM *must* be done in Common Ground. Common Ground will record the merge in the merge log, and the Connector will read that information and perform the same merge in COM the next time the sync is run. If a user attempts to merge such records in COM, a pop-up warning will prompt the user to perform the merge in Common Ground instead. Since the records are in "quarantine" in Resolve New Registrations in COM, performing the merge in Common Ground will release them from the quarantine.

If two records are merged in Common Ground and only one of them is in COM, the Connector will automatically update the record's IDs, if necessary.

D. Other Automated Data Integration

There are many reasons why an organization may have a custom solution to exchange data with or import data from another database. Some examples are lockbox files, mail house address updates, telemarketing result files, records from a program database, integration with an accounting system, etc. Because there are so many possible uses and specifics, it is beyond the scope of this guide to go into all the best practices associated with setting up a custom import or integration.

In terms of data hygiene, if at all possible the other system should store the Common Ground/Salesforce Contact ID. This will make it clear which records are new versus existing when processing a file. As a second step, it is ideal if the import/integration process checks for duplicates when it thinks it has a new record. The record may be new to the other system, but may be in Common Ground already.

III. Cleaning Existing Common Ground Data

Stopping the stream of duplicates into your Common Ground system with knowledge of native duplicate-checking features, how use the available import tools, and Common Ground-COM Connector features is half the battle. Another aspect of data hygiene is managing duplicates that have already snuck in.

There will always be at least a few duplicates that get into your Common Ground system. By following the recommendations in this section, however, they will not be there for long!

A. Common Ground Duplicate Management

Common Ground offers a duplicate resolution tool on the Utilities tab called Duplicate Management. This tool is designed to find and display duplicate Contacts based on a user-specified criterion (such as Full Name or Email Address) or a more complex pre-set or custom "match key" (combination of whole or partial field values). Once potential duplicates are displayed, a user goes through each group to determine whether or not the records really are duplicates, and if so, which information should be taken from each record in order to build the best surviving record. This is a record-by-record, field-by-field process. More information about this tool is available in the Resource Center: https://secure2.convio.net/customer/commonground/v25/guides/Duplicate_Management.htm

Using the Common Ground Duplicate Management tool is the recommended way to clean up duplicate Contacts because it is designed to deal with all the specialized, non-native Salesforce objects and fields that make up much of the Common Ground functionality that fundraisers need. In particular, Relationships, Tributes, Donations, Pledges and Recurring Gifts have special significance for duplicate management, and the Common Ground Duplicate Management function is built to handle these pieces of data. Other available tools do not always take these CG-specific parts of the record into account — more on that follows.

Some terminology: When merging Contacts, the one that is going to be kept is called the *master*. The record that is going to be removed from the system is called the *servant*. It's important for your organization to develop criteria so that users will know how to identify the best master record. Some popular criteria are:

- Has gifts
- Older created date
- Has an email address

NOTE: It is a known issue that records selected as "Not Duplicate" will show up again in future duplicate searches. This is planned to be resolved in the Fall 2010 release of Common Ground. In the meantime, one workaround is to use the "Only show Contacts created since" option to narrow the search to records created only since a specified date (such as the last date your system was examined for duplicates). The tool will then compare only the new records to all records within the system.

There are some things to be aware of when merging records, so we have prepared this situational guide that will help your users resolve duplicates most effectively, with the fewest number of steps.

1. **Once two Contacts are merged, if there were no other Contacts on the servant Account, an empty Account record (sometimes called a "ghost" account) will be left behind.** The Fall 2010 release of Common Ground is predicted to resolve "ghost accounts" so that the empty Account will be automatically removed.

   However, if there are additional Contacts remaining on the Account, you will need to deal with them before the Account can be removed. Usually, these Contacts need to be merged into the same Account as the one just affected by the Contact merge. The best way to accomplish this is to open a separate tab and look up the Account for each Contact that was just deleted as part of a merge. You should do this as you go through the Contact merge process, and resolve the Accounts before moving on to the next Contact merge.
   - If there are no additional Contacts, determine if there is any information on the Account that should be moved to the surviving Account (addresses, relationships), and if so, move it. If there is no useful information, delete the Account.
   - If there are additional Contacts, determine if those should be moved to the other Account or not, and if so, move them. If doing so results in an Account with no Contacts left in it, proceed with the step above.

2. **You can tell if a Contact is an Organization Contact by the name of the associated Account.** If someone is an Organization Contact and also has their own Household Account (with donations), it is best to merge the Organization Contact into the Household Contact. Then, follow up on the Contact record and add the relationship to the Organization Account. If the Household Contact doesn't have any compelling information, and you don't really know the person other than as a representative of the organization, it might be better to merge the Household Contact into the Organization Contact and remove the extraneous Household Account.

3. **Duplicate Donor Proxy Contacts will appear in the Duplicate Management tool.** They will be recognizable because the Organization name will be the

Contact name. Recall that Donor Proxy Contacts cannot be deleted. It is best to save these merges until last since you will have to go into the servant Contact to change the record type before you can proceed with the merge.

On duplicate Organization Accounts, if at least one of the Donor Proxy Contacts has any of the key information (Relationships, Tributes, Pledges, Recurring Gifts, Donations), the first step is to change the contact record type of the servant donor proxy to Organization Contact. Also, recall scenario #1 about checking for additional Contacts besides the Donor Proxies.

Then use the Common Ground merge tool to merge the contact records. If there are no more Contacts on the servant Account, that Account can be merged into the master Account. If there are additional Contacts to be merged, follow the guidelines in scenario #1 for dealing with them.

4. **On duplicate Organization Accounts in which none of the Contacts (not just Donor Proxies) has any of the key information (Relationships, Tributes, Pledges, Recurring Gifts, Donations), begin by changing the record type on the donor proxy contact of the servant Account to Organization Contact.** Then, merge the Accounts (in the Tools area of the Accounts page). If there is no other information on the servant donor proxy contact, it can be deleted. If there is additional information, follow the procedure in scenario 3 for merging donor proxies. If there are additional contacts besides the Donor Proxy contacts, use the Duplicate Management tool to consolidate them.

B. CRMfusion Tools

In the previous section about importing, we mentioned a CRMfusion tool called DupeBlocker. This app (available in the Salesforce.com AppExchange) stops users from being able to enter a duplicate record. Once installed, an administrator can use a default scenario or develop a custom scenario to identify duplicates. These scenarios run every time a user enters or edits a record. The administrator can decide whether to block the user from saving the duplicate, or just put up a warning message.

DemandTools has a function for merging duplicate Contacts (both singly and en masse), but in most cases it will not be the best choice for Common Ground users. Here's why:

- The presence of Donations, Pledges, Recurring Gifts and Tributes will stop the merge from completing. (The Common Ground tool can merge these pieces of data.)
- Additionally, if the servant Contact has Relationships, the DemandTools duplicate merge function will complete the merge, but will drop the Relationship

information and the information will be lost. (The Common Ground tool does not lose this information.)

- For organizations using the Convio Common Ground Connector, there are additional considerations:
    - DemandTools deletes and recreates TeamRaiser™ and Tributes Registrations, which can create errors and data problems for the Connector.
    - The DemandTools merge function does not populate the Common Ground merge log, which is what synchronizes merges with COM.

For these reasons, we advise Common Ground users to use the Common Ground Duplicate Management tool. It's possible that non-fundraising-related records in your Common Ground system that do not use any of these susceptible areas could use the DemandTools functionality.

IV. Assessing Your System

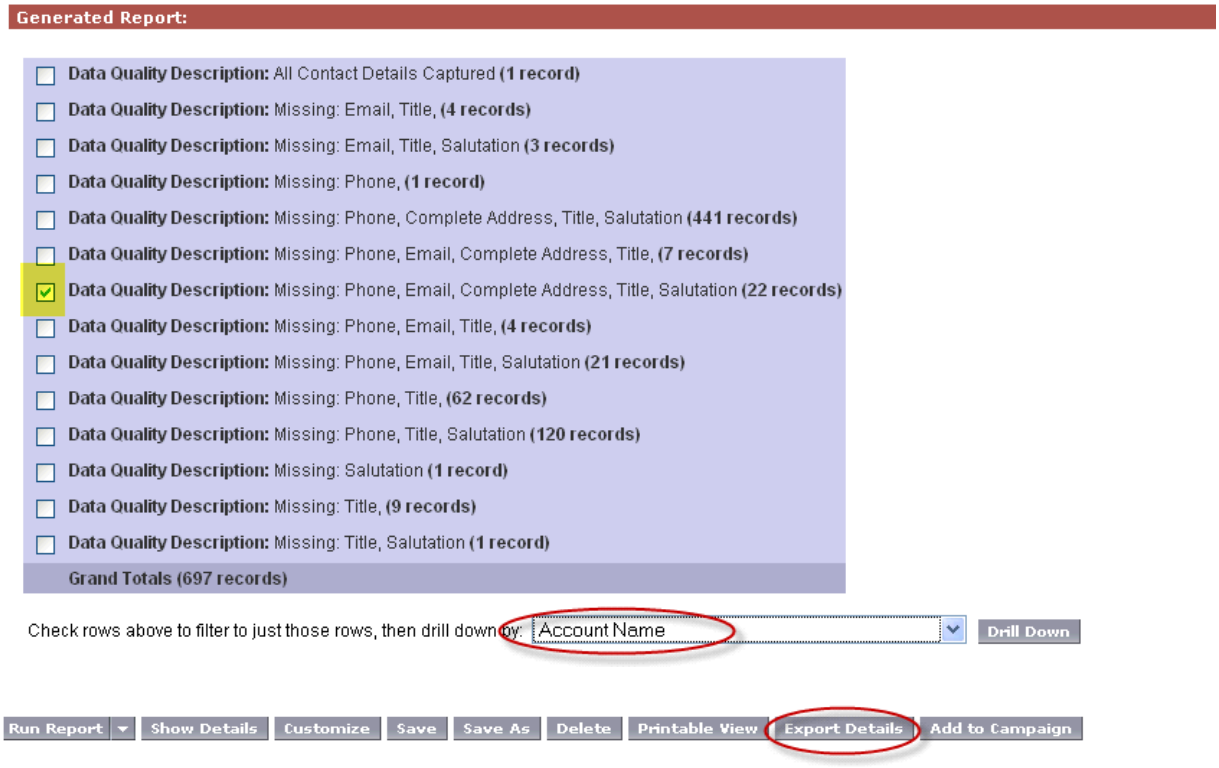
A. How to Assess Your Data Quality

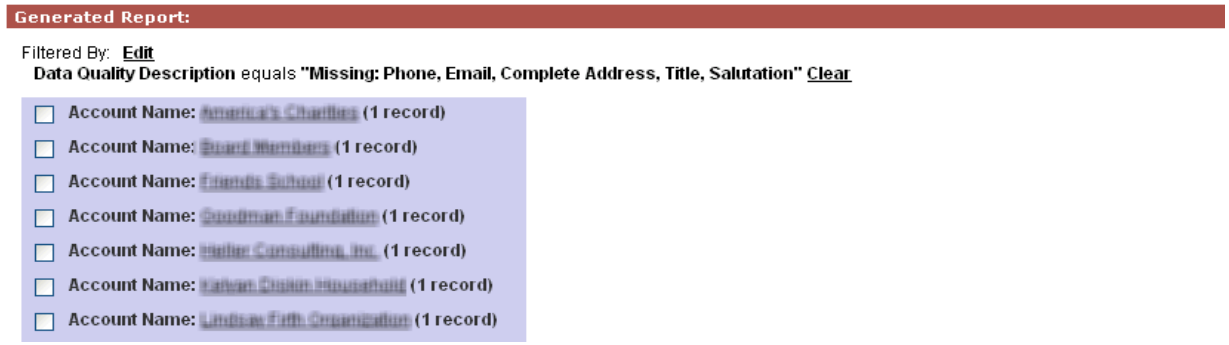
Database administrators often wonder how good their data is. For example, in a database of 80,000 records, it's not uncommon for only 60,000 (or fewer) to be contactable. What is going on with the rest of those records?

It could be that those records simply opted out of communications from your organization, or are deceased individuals, or don't have good address/email information. Or, it could be that these records are missing a key piece of information (or two) that keeps them out of your communications. Perhaps some of these records are duplicates or don't have a minimum of useful information.

One way to start getting a handle on your data quality is to install the Data Quality Analysis toolset (in the AppExchange as Data Quality Analysis Dashboards by Force.com Labs). This application provides dashboards and reports to help you identify records that are missing key pieces of information. The most helpful reports for Common Ground administrators are likely to be the Account, Contact and Opportunity reports. There are three reports for each object: Data Quality Descriptions, Data Quality Score, and Owner Data Quality.

When the reports run, they summarize the number of results by the different criteria. The example below is of the Contact Data Quality Description report. Once this summary is available, you can select an area to see more details about which records are affected.





After drilling down to see the list of affected Accounts, you can click on the hyperlinks to open the Account record. From there you can open the Contacts and review them.

Consider updating the date range each time the report is run, so that records that have previously been reviewed do not appear again.

If staffing resources permit, this is potentially a good task for data entry operators to take over. They can be trained to run the reports, or the drill-down can be exported to Excel and sent to them to review the records.

Going through this audit process is a good way to become aware of areas where data quality issues can be resolved by making a field required, or by implementing a

validation rule. For example, a validation rule could prompt a user to enter a ZIP code if an Address and City are entered.

Once armed with the knowledge of where your most common data quality pitfalls lie, you can help your users to either complete records more thoroughly, or properly mark records as un-contactable.

The Data Quality Analysis toolset also comes with several dashboards. These can be accessed on the Dashboards tab or on the Home Page. More information is provided on the Dashboards tab than on the Home Page. The information provided on the Home Page is a more high-level snapshot of data quality.

If your organization is comfortable with using formulas, it is possible to modify the formulas used for these reports and dashboards in order to more precisely target the fields your organization considers important on records. The formulas are located in Setup → App Setup → Customize → Contacts → Fields. The fields are custom Contact fields: Data Quality Description and Data Quality Score. There are two things that are important to note about these formulas: 1) If you change one, you should also change the other, and 2) Only modify these formulas if you have experience with formulas. Modifying formulas is an advanced administrator skill.

B. Do We Have a Duplicate Problem?

We are often asked by our clients if we think a database has a duplicate "problem." It is important to remember that the number of potential duplicates depends entirely on the criteria used. If your criteria are loose, your system will appear to have a greater number of duplicates than perhaps it actually does. It's a good idea to determine a reasonable set of criteria for potential duplicates and use it consistently.

In our experience, a database with less than 3% duplicate records does not have a duplicate problem. For example, in a database of 80,000 records, this would be 2,400 potential duplicates or fewer. In a system like this, it is relatively unlikely that duplicates are causing problems with duplicate mailings or with donor relations. This is because oftentimes a system in this situation has a number of records that are un-contactable, and the potential duplicates either counteract these records or are also un-contactable.

As a system rises above a 5% duplicate threshold, it becomes more likely that duplicates will start to become the source of mailing or donor relations problems. We have seen systems that were as high as 25-30% potential duplicate records. In a situation like this, data hygiene becomes of paramount importance.

There are a couple of ways to figure out how many total potential duplicates are in your system. DemandTools will give the total number of Contacts and the total number of

groups of potential duplicates, after the duplicate scenario is established. The Common Ground Duplicate Management utility gives the total number of groups (rows on the screen), but not the total number of Contacts.

C. Putting Practices into Place

One of the most valuable projects a Common Ground administrator can undertake is a comprehensive data hygiene internal guide. Ideally, an internal data hygiene guide will both lay out organizational policy around data hygiene, and also provide instruction on how to manage duplicates and assess data quality.

Here are some examples of topics that should be included in a data hygiene guide:

- Instructions for entering Contacts, Households and Organizations, including required fields.
- Instructions for making common updates to records. We call these "What to Do When." For example:
    - Changing an address
    - Marking a donor as deceased
    - Changing an employer
    - Adding a spouse
    - Marking a bad address
    - Changing contact preferences
- Circumstances and procedures for promoting an Organization Contact to a Household Contact.
- Procedures for regularly-run imports.
- Instructions for identifying a master record.
- Procedures for merging records (akin to the scenarios mentioned in this document), including dealing with Account records.
- Procedures for data quality reports to run and steps to take with the results.
- Documentation of data quality-related validation rules.

You can use this list as a starting point for the topics that should be covered in your organization's data hygiene guide. It's easy to make documentation such as this available to your users via the Documents tab.