

Maryland Hillel Data Architecture Review

Introduction

Maryland Hillel has been experiencing inconsistencies in their student data that has been hindering their operations. An example of these issues are multiple records for the same student with different contact or personal information. Or incorrect relationship links between family members. This is an issue because it slows down operations, potentially turns away potential donors or new members, and incur additional costs in time and resources in order to identify and correct.

Maryland Hillel sponsors a diverse range of activities including social, cultural, religious, and educational events and programs. Maryland Hillel organizes events weekly, including providing Kosher meals daily which follow Jewish dietary restrictions. Their contact for this project is Shawn Laing, director of finance and operations. and will be acting as the main moderator and the client.

Maryland Hillel gets their data from multiple sources. They receive some student information directly from the University of Maryland and some from Hillel International, but Maryland Hillel primarily gets their information from their members directly upon registration. They also record information when they receive phone calls or email, through forms at events, or through chance in-person encounters. With so many data entry points and without strict data governance policies, it's easy to see how inconsistencies in the data may occur.

Maryland Hillel has contacted students in the University of Maryland's MIM Capstone course to review their data architecture in order to identify the origins of these issues and identify potential tools that can prevent, detect, and correct these inconsistencies in the data. The project team will review Maryland Hillel's processes for record collection, entry, update, and use in order to give recommendations on potential solutions.

Overview

As Maryland Hillel primarily serves students at the University of Maryland, the data they collect involves students, prospective students, alumni, and parents. They collect the basic contact info such as name, address, email, and phone number, but they also collect more detailed student information such as graduating year, your campus address, whether or not you are on honor roll, and whether or not you are in Freshman Connection, Greek life, and ROTC. Maryland Hillel also tracks event and donation history, in a yearly and lifetime (since 2019) capacity.

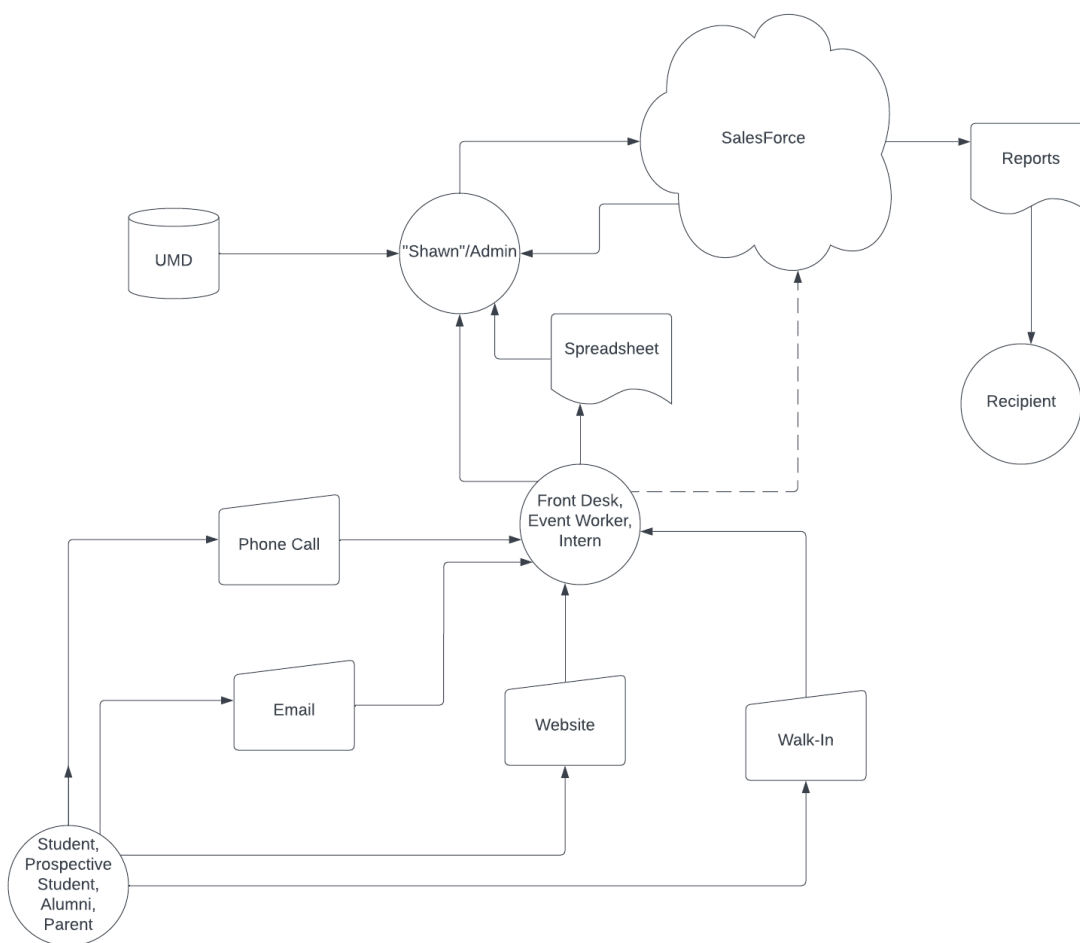
Maryland Hillel uses Salesforce as their Custom Relationship Management system. Student, alumni, and parent information is entered manually. There are two main tables, Households, which tracks students, alumni, and parent personal information, and Campaigns,

Maryland Hillel Data Architecture Review

which tracks events and donations. Name, Address, and Email are typically used as Primary Keys. As of now, Maryland Hillel's database contains over 100,000 records and grows by roughly 1,500 records per year, with 500-600 new students and members each semester.

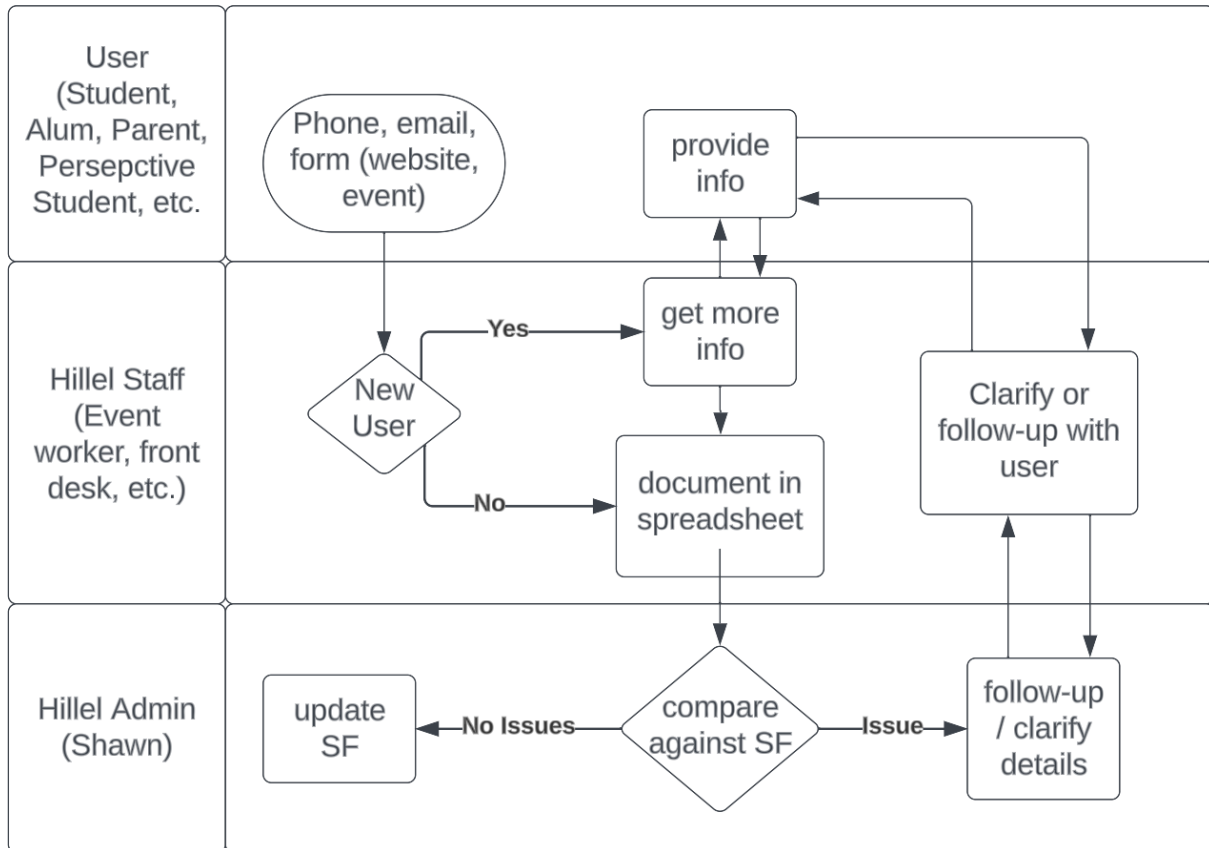
There are three main data entry points: phone, email, and forms on their website or at in-person events. Maryland Hillel also receives student data from the University of Maryland systems. Furthermore, records may be supplemented by knowledge from personal relationships or outside knowledge. The current process looks something like this: when Maryland Hillel is contacted, their information is taken down and emailed to someone to put into Salesforce. This is supposed to be Shawn or the database admin, but occasionally someone else in the organization will upload the data themselves.

Data Flow Diagram: Shows how information moves through Maryland Hillel's system



The primary purpose of this data is to create reports, with the most frequently generated reports being mailing lists. The targets of these mailing lists can vary depending on the purpose, depending on if Maryland Hillel is trying to spread information or make a request. Some examples are students, freshmen, seniors, alumni, recent grads, parents, etc.

Swimlane Diagram: Indicates Data Touch points



Reports are typically generated 2-4 times a month and this is usually how inconsistencies are detected. Other data issues are occasionally found by chance, such as when entering or looking up a record. Hillel does use Salesforce's process for checking for duplicates, but it doesn't detect all of them. Currently, it is set to detect for first name, last name, and email. When those reports are generated every few weeks, roughly 15-20 inconsistencies are detected.

Challenges

The Hillel team is currently grappling with several data inconsistencies in their Salesforce database, which can potentially cause operational challenges and impact donor relationships. Hillel's primary inconsistencies are Duplicate Entries, Missing and Incorrect Key Information, and Incorrect Graduation Year and Phone Numbers.

These challenges are presenting themselves because there's an overall issue with the delegation of roles and responsibilities. Multiple people are entering data into the system simultaneously so there are a lot of duplicate records. These entries do not go through a check before entering so these records may already exist in the system.

The ramifications of these challenges extend to potential donors, impacting their willingness to contribute and potentially causing frustration due to recurrent emails and communications from Hillel. The need for a more structured and coordinated approach to data entry is evident, emphasizing the importance of implementing measures to mitigate duplicate entries and ensure data accuracy for improved donor engagement.

Moreover, the adverse impact extends to the broader community, encompassing students, alumni, donors, and parents. In the absence of a meticulous data entry method, there is a heightened probability that these key stakeholders may overlook essential information. A structured and systematic approach to data entry is imperative to mitigate these challenges, ensuring a seamless communication flow and minimizing the risk of donor disengagement and stakeholder information lapses.

Duplicate Entries

Duplicate entries in the Salesforce database present a significant problem. When multiple entries for the same individual exist, the system may generate and send redundant reports to the same person. This not only leads to inefficiencies in resource allocation but can also be quite bothersome for the recipients. Annoyed individuals are less likely to engage or donate.

Missing and Incorrect Key Information

The database may contain entries with missing or inaccurate key information, such as names, email addresses, or physical addresses. These inaccuracies often arise when individuals provide incorrect email addresses or, in some cases, personal email addresses that are not associated with their engagement. Furthermore, data discrepancies may include instances where first names and last names are swapped, leading to confusion. Another critical issue is when students, alumni, parents, or donors have multiple addresses within the system. This makes it difficult to determine their primary contact information, causing individuals to miss out on essential communications.

Incorrect Graduation Year and Phone Numbers

It's not uncommon for incoming freshmen to provide their anticipated graduation year when they initially express interest in Hillel. However, there are instances where these graduation years change as students' progress through their academic journey, and this information is not consistently updated in Salesforce. Similarly, discrepancies in phone numbers can lead to communication challenges. Accurate phone numbers are vital for staying in touch with students, alumni, and donors. Any inconsistencies in this data can hamper the ability to reach out effectively.

Team Roles

1. Main Moderator
2. Data Administrators
3. Receptionist

4. Student Interns
5. Data Analyst (new role)

Main Moderator (Shawn) and One Other Administrator

The primary responsibility of the Main Moderator, Shawn, and one other administrator is to maintain and update the data within the Salesforce database. These individuals are granted elevated access privileges to make changes to the data stored in Salesforce. Their role includes ensuring that the database is accurate, up-to-date, and well-organized. This critical role requires a deep understanding of the data management processes within Hillel.

Data Administrators

Data administrators play a pivotal role in managing and accessing Salesforce but do not have permission to make direct changes to the data. Instead, their responsibilities primarily revolve around data retrieval, analysis, and support functions. They assist the Main Moderator and administrator by providing insights and generating reports based on the existing data. This collaborative approach ensures that data remains secure and accurate while allowing for efficient data utilization.

Receptionist

The receptionist serves as the initial point of contact for students, parents, alumni, or donors seeking information about Hillel. When someone expresses interest and requests information, the receptionist is responsible for providing them with the required forms. These forms are typically used to collect essential information about the individual, ensuring that accurate and relevant data is entered into the Salesforce system.

Student Interns

Student interns play a supportive role in the overall data management process. While their primary responsibilities may involve other administrative tasks within Hillel, they can contribute to data-related activities as needed. This can include tasks such as data entry, document management, or assisting in the collection and organization of data. Student interns help ensure that the data management workflow operates smoothly and efficiently.

Data Analyst

The data analyst is a role that can be its own role or assigned to one of the above roles. The data analyst is responsible for handling the data inconsistencies, by running tools to identify any and applying deduplication procedures to correct them. Assigning these data inconsistencies to someone will reduce their occurrences during reports and other critical functions.

Improving Data

There are several ways to improve the quality of Maryland Hillel's data. Here are some suggestions that can be implemented in order to reduce the inconsistencies in the data.

Limitation of Access

- Access, as a database management system, comes with certain limitations when it comes to managing user access to specific records. There should be a system implemented which allows for certain members of the team to access certain records.
- A chart can be presented to visualize which members have access to which systems to avoid repetition in the data collection

Plan:

- Survey the team to see who has access to what systems
- Analyze potential discrepancies
- Create a visual plan with each role
- Delegate each role with specific access requirements
- Train staff on new standards

New Preliminary Spreadsheet

- Any point in which data in SF would need to be changed needs to be documented in this spreadsheet
- Registration, information seeking, donating, new users, updating information, etc.
- This spreadsheet would act as a safety net to check against before adding a potential duplicate into SF
- The spreadsheet would include the basic contact info, purpose of the inquiry, and whether or not they might be a current student.
- This is to reduce the amount of data that would need to be checked against SF and help guide the admin as to what would need to be changed in SF

Plan:

- Plan structure of spreadsheet
 - Determine the columns needed for spreadsheet such as current student status, Basic info(name, email, phone)
- Create spreadsheet template on Excel
- Create guidelines for spreadsheet like what data goes in what column
 - Documentation Process- Define when and how the spreadsheet should be updated
- Cross-Reference and Check
 - Ensure that the spreadsheet is cross-referenced regularly with SF
- Set schedule to periodically review spreadsheet

Data Cleansing

- Data Cleaning is done to reduce discrepancies in data and eliminate duplicates
- Set out a weekly or biweekly schedule for data-cleaning processes
- Perform an audit
 - Regularly perform a data audit to identify discrepancies in the data. This can also be done on a weekly or biweekly schedule
- Document Data entry points
- Search for duplicates, missing data etc.
 - Third party data cleaning tools can be utilized for this practice. There are many different options for these tools such as DemandTools or OpenRefine.

Plan:

- Define Data Cleaning Processes
 - Define the scope of data cleaning—what data sources, fields, or systems need cleansing.
 - Identify and document all data entry points and sources contributing to your datasets.
- Establish Cleaning Schedule
 - Designate a weekly or biweekly schedule for data cleaning processes:
- Data Audit Implementation
 - Define criteria for discrepancies, duplicates, and missing data.
 - Create checklists or guidelines for auditing data consistency and quality.
 - Frequency: Perform audits on a regular basis (weekly or biweekly) to identify and rectify discrepancies promptly.
- Data Cleaning Techniques
 - Utilize third-party data cleaning tools like DemandTools or OpenRefine to identify and eliminate duplicates.
- Create a step-by-step guide or checklist for executing data cleaning tasks for designated roles
- Verify that data cleaning processes align with existing data management system (Salesforce)

Implement Ticket System

- Utilize the project management tool JIRA
- This will allow for ticket creation, so at meetings new tickets can be created to deal with each issue
- A scrum-style process can be implemented to go through each ticket on a weekly basis
- The tool will assign each task to different roles and create an overall structure for dealing with new issues

Plan:

- Identify and outline the needs for the team: managing data entry
- A team member should configure JIRA to create workflows
- Train the team on how to use the platform effectively
- Define a workflow style such as Scrum for ticket implementation
- Define guidelines for ticket creation and set up weekly or biweekly meetings to go through tickets

Implement Training Documentation

- To ensure that the practices are standard and only people of certain roles have certain access privileges, there should be a proper documentation that should be used for training purposes
- This can be an SOP which outlines the duties of each member
- The training will specify what information is being put in by what staff and the standard data practices

Plan:

- Identify roles within the organization that interact with data and systems. Define the access privileges associated with each role.
- Develop detailed Standard Operating Procedures (SOPs) for each role:
 - Outline specific duties, responsibilities, and tasks associated with data handling and access privileges.
 - Include guidelines on data entry, data security practices, and compliance regulations.
- Create comprehensive training materials based on the SOPs:
 - Develop training manuals, guides, or presentations that align with the SOPs.
 - Incorporate practical examples and scenarios to illustrate standard data practices.
- Integrate the training documentation into the onboarding process for new hires.

Data Validation Procedure

- Provide rules that will eliminate incorrect data to be entered into the system
- Establish rules for date format, length, range, uniqueness, or relationship between fields.
- Examples of Rules:
 - Formatting rules: Make sure that email address is in proper format, establishing phone number formats
 - Length rules: Length of UID or phone numbers
 - Relationship Rules: validate that one field correlates with other fields such as parent and child information or student and UID information.
 - Uniqueness Rules: ensures that an item is not entered multiple times into a database such as making sure each UID is unique

Recommendations fix

1. Data Validation Procedure

Impact: High

Reason: Establishing clear rules and validation procedures ensures that only accurate and relevant data is entered into the system, significantly reducing errors and maintaining data integrity. Having a clear documentation of rules will be helpful for the team to refer to when managing data or entering data into Salesforce.

2. Implement Ticket System

Impact: Medium to High

Reason: A ticket system helps in tracking and resolving data issues efficiently. Assigning tasks, managing priorities, and ensuring accountability streamline data management processes.

3. Data Cleansing

Impact: Medium

Reason: Regular data cleaning and audits are essential for maintaining data accuracy. However, while important, it might be slightly less impactful compared to establishing clear validation rules initially.

4. Limitation of Access

Impact: Medium

Reason: Managing user access is crucial, but it might not directly impact data quality as much as other measures. However, controlling who can access and modify data is vital for security and accuracy.

5. Training Documentation

Impact: Medium to Low

Reason: Training is important, but it might have a lower immediate impact compared to other measures. Nevertheless, ensuring standardized practices and access privileges is crucial for long-term data quality.

6. New Preliminary Spreadsheet

Impact: Medium

Reason: While creating a safety net for potential duplicates is beneficial, heavily relying on external spreadsheets might add complexity and room for error. It's essential but might have a lower impact compared to other measures focused on direct data management within Salesforce.

Data Deduplication Tools

Below are 6 Data Deduplication tools that we have found and researched and how they might be used to solve data inconsistency issues.

DemandTools (<https://www.validity.com/demandtools/pricing/>)

DemandTools is a tool that allows organizations to clean and manage Salesforce data. It integrates directly with Salesforce and has the following features for resolving data inconsistencies.

1. **Data Standardization:** The tool enables users to standardize data within Salesforce. It can normalize and cleanse data by correcting formatting inconsistencies, ensuring uniformity across records.
2. **Bulk Data Cleansing:** DemandTools supports bulk data cleansing operations. Users can execute mass updates to correct inconsistencies or errors across multiple records simultaneously. This helps in rectifying data inconsistencies in a time-efficient manner.
3. **Advanced Matching Algorithms:** Its advanced matching algorithms assist in identifying similar records that might not be exact duplicates but exhibit inconsistencies. This helps in grouping records for further review and resolution of inconsistencies.

Maryland Hillel Data Architecture Review

4. **Scheduled Jobs for Continuous Maintenance:** Users can schedule jobs within DemandTools to run at regular intervals. These scheduled jobs can perform routine data cleansing, deduplication, and consistency checks to maintain data integrity over time.
5. **Customizable Data Quality Checks:** The platform allows users to define custom rules and criteria to identify and flag data inconsistencies. This customization ensures that specific data quality issues are addressed according to the organization's requirements.
6. **Reporting and Data Analysis:** DemandTools offers reporting functionalities to visualize data inconsistencies. Users can generate reports to track changes, monitor data quality metrics, and identify trends in data inconsistencies for further analysis and improvement.

Pros	Cons
<ul style="list-style-type: none"> - modern interface - automated data mangement jobs - can schedule jobs and scenarios - can define "duplicate" (exact, relaxed, cleaned) 	<ul style="list-style-type: none"> - monthly cost (\$5.55/\$11.00 per month)

DemandTools has 3 plans, including a free version that includes a 2-week trial for the paid deduplication features. All 3 plans include live, weekly office hours, email support, and documentation and video libraries. The full version (\$11/mo) includes a success manager for personalized on-boarding.

The screenshot displays the DemandTools web application. At the top, there's a navigation bar with the DemandTools logo and user information. Below this, a 'Start a new scenario' section contains buttons for Convert, Dedupe, Delete, Export, Import, Match, Modify, Reassign, Tune, and Verify. To the right, an 'Assess' button is accompanied by a 'Record Quality Bar (as of 2/17/23)' showing a green bar. Below these, a 'Record Activity for past 30 days: 331 Scenario Runs' summary shows statistics: 252,763 Records Merged, 2,964 Records Inserted, 203,107 Records Updated, and 149,925 Records Modified. The main section is titled 'Scenarios' and includes a search bar and filters for Module, Object, Created By, and Tags. A table lists various scenarios with columns for Last Run, Scenario, Module, Object, Created By, Runs, and Schedule. The scenarios include 'Account Duplicates Cleaned Name', 'Account Name Match', 'Account Name Street Address', 'Account Name Street Address - Prospect', and several 'Accounts(BillingCountry) - ISO 2Character' and 'Accounts(BillingCountry) - ISO 3Character' entries.

LAST RUN	SCENARIO	MODULE	OBJECT	CREATED BY	RUNS	SCHEDULE
never	Account Duplicates Cleaned Name	Dedupe	Account	Stuart Smith	0	:
a year ago	Account Name Match	Match	Account	Stuart Smith	12	:
4 months ago	Account Name Street Address	Dedupe	Account	Stuart Smith	12	:
14 hours ago	Account Name Street Address - Prospect	Dedupe	Account	Stuart Smith	112	:
a year ago	Accounts(BillingCountry) - ISO 2Character	Modify	Account	Validity	4	:
never	Accounts(BillingCountry) - ISO 3Character	Modify	Account	Validity	0	:
never	Accounts(BillingCountry) - LongName	Modify	Account	Validity	0	:
never	Accounts(BillingCountry) - US Filler	Modify	Account	Validity	0	:
never	Accounts(Billing) - Loose Criteria	Dedupe	Account	Validity	0	:
a year ago	Accounts(Billing) - Rigid Criteria - Recommended First Pass	Dedupe	Account	Validity	1	:
a year ago	Accounts(Billing) - Semi Rigid Criteria	Dedupe	Account	Validity	2	:

OpenRefine (<https://openrefine.org/docs/manual/cellediting#cluster-and-edit>)

OpenRefine is a powerful and free, open-source tool for working with messy data. OpenRefine allows mass cleaning, transformation, and normalization of data. It is an external

application so data from Salesforce would have to be downloaded (as a csv, excel file, etc.). Here are some ways in which OpenRefine can be used to identify and correct data inconsistencies.

1. **Facilitating Data Matching:** OpenRefine can assist in matching and merging duplicate records by using powerful clustering algorithms. It identifies similar or identical records and groups them together for review and consolidation. This helps in identifying potential duplicates within the student dataset.
2. **Faceted Browsing and Exploration:** The tool offers faceted browsing, enabling users to explore and filter data based on different facets or characteristics. This functionality helps in identifying patterns, outliers, and inconsistencies within the dataset.
3. **Bulk Editing and Transformation:** OpenRefine allows for bulk editing and transformation of data. Users can perform operations on large datasets efficiently, making it easier to correct errors and inconsistencies across multiple records simultaneously.
4. **Customized Data Transformation:** It provides a range of transformation functions and scripting capabilities that allow users to define custom operations to clean, transform, and reconcile data based on specific requirements. This can be particularly useful for resolving complex data issues.
5. **Export and Integration:** OpenRefine enables the export of cleaned and standardized data into various formats compatible with other systems or databases, facilitating integration with existing tools or platforms used by University of Maryland Hillel.
6. **Auditing and Reconciliation:** It keeps a record of all operations performed on the data, facilitating auditing, and ensuring the traceability of changes. This feature is valuable for maintaining data integrity during the cleaning and deduplication process.

Pros	Cons
<ul style="list-style-type: none"> - free - privacy: data is cleaned locally, not externally located - utilize facets and apply operations to filter views - fix inconsistencies by clustering similar values - reconcile the data by matching with external databases - infinite undo/redo: rewind and replay any previous states and operation history 	<ul style="list-style-type: none"> - data is cleaned locally - not as intuitive to use - not integrated into Salesforce - interface more similar to spreadsheet format

Although OpenRefine may be used for deduplication, the program is not necessarily designed with Salesforce in mind. It is a powerful tool often used when cleaning large amounts of data in spreadsheets with features that may be useful.

Cluster & Edit column "Categories"

Multiple cluster methods with multiple key/distance functions detect and solve several variations.

Method: key collision Keying Function: ngram-fingerprint Ngram Size: 3

Cluster Size	Row Count	Values in Cluster	Merge?	New Cell Value
2	9	<ul style="list-style-type: none"> Air bricks (8 rows) Airbricks (1 rows) 	<input checked="" type="checkbox"/>	Air bricks
2	351	<ul style="list-style-type: none"> Transport-Water (350 rows) Transport - Water (1 rows) 	<input checked="" type="checkbox"/>	Transport-Water
2	8	<ul style="list-style-type: none"> Doorknobs (7 rows) Door knobs (1 rows) 	<input checked="" type="checkbox"/>	Doorknobs
2	2	<ul style="list-style-type: none"> Band saws (1 rows) Bandsaws (1 rows) 	<input checked="" type="checkbox"/>	Band saws
2	12	<ul style="list-style-type: none"> Bookmarks (11 rows) book marks (1 rows) 	<input checked="" type="checkbox"/>	Bookmarks
2	261	<ul style="list-style-type: none"> Swatch books (207 rows) Swatchbooks (54 rows) 	<input checked="" type="checkbox"/>	Swatch books
2	11	<ul style="list-style-type: none"> Mailbags (8 rows) Mail bags (3 rows) 	<input checked="" type="checkbox"/>	Mailbags
2	4	<ul style="list-style-type: none"> Skullcaps (3 rows) Skull caps (1 rows) 	<input checked="" type="checkbox"/>	Skullcaps

spacing differences

capitalization differences

spelling differences

11 clusters found

Rows in Cluster: 0 — 680

Average Length of Choices: 8.5 — 16

Length Variance of Choices: 0.5 — 1

effect of the selected cluster method on the entire dataset

Select All Deselect All Merge Selected & Re-Cluster Merge Selected & Close Close

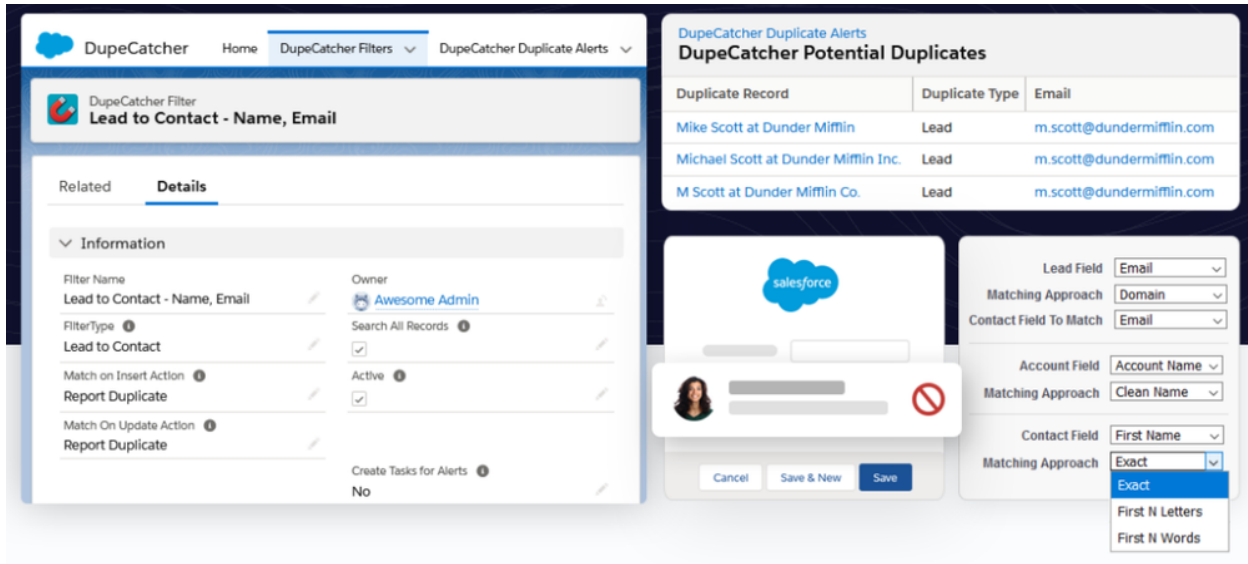
DupeCatcher (<https://dupecatcher.com/#features>)

DupeCatcher is a free data deduplication tool designed with Salesforce. It was developed by Symphonic Source, Inc. DupeCatcher aims to identify and eliminate duplicate records within Salesforce, ensuring data accuracy and consistency. The tool provides features such as customizable matching rules, real-time deduplication, and merging or deleting duplicate records. Some of the ways DupeCatcher can find and rectify duplicates

1. **Creating Rules and Filters:** DupeCatcher functions based on Filters and Rules are set by a Salesforce administrator. Once installed, admins will see a DupeCatcher option in the Salesforce dropdown menu. Admins can create new filters and rules based on their requirements to identify duplicates.
2. **Blocking Duplicates:** DupeCatcher can block a duplicate record from being created and display a warning and table with a list of hyperlinks to matched records. DupeCatcher has an override function which allows certain individual business processes and policies to require duplicate records.
3. **Merge Function:** DupeCatcher makes it even easier to manage the data entering your Salesforce with its Merge function. The Merge feature allows users to combine identified duplicates into a single record of the same type.

Maryland Hillel Data Architecture Review

Pros	Cons
<ul style="list-style-type: none"> - easy set-up - user-friendly - built-in app in Salesforce - free 	<ul style="list-style-type: none"> - older interface - little-to-few customization options - little-to-few features



Datablist (<https://www.datablist.com/features/duplicates-remover>)

Datablist is a free cloud based platform offering data cleansing solutions. It offers a data and automation platform to move processes to algorithms by aggregating data in a single place and starting building applications. It also cleans, transforms, and organizes sales, customers, or product data. This is an external application; the data should be uploaded as CSV to identify duplicates.

1. **Selected Properties Mode:** This mode likely enables users to focus on specific properties, such as email addresses for contacts or website URLs for companies. Users can choose to identify duplicates based on these selected properties, streamlining the deduplication process.
2. **Merging Duplicates with Identical Property Values:** When duplicate items have identical property values across all fields, Datablist automatically retains one item and deletes the others. This ensures that no information is lost, as all duplicates contain the same data.
3. **Conflict Resolution for Duplicates with Conflicting Property Values:** When duplicate items have conflicting property values (i.e., different information for the same field), Datablist identifies these conflicts. Instead of automatically merging conflicting items, Datablist skips them and leaves them for manual review and merging by the user. This allows users to make informed decisions about how to resolve conflicts, ensuring accuracy in cases where the system cannot determine the correct values automatically.
4. **Automatic Merge during Imports:** In addition to the Duplicates Finder, you can set a 'Unique values' setting on your collection properties. During data import, when an item

Maryland Hillel Data Architecture Review

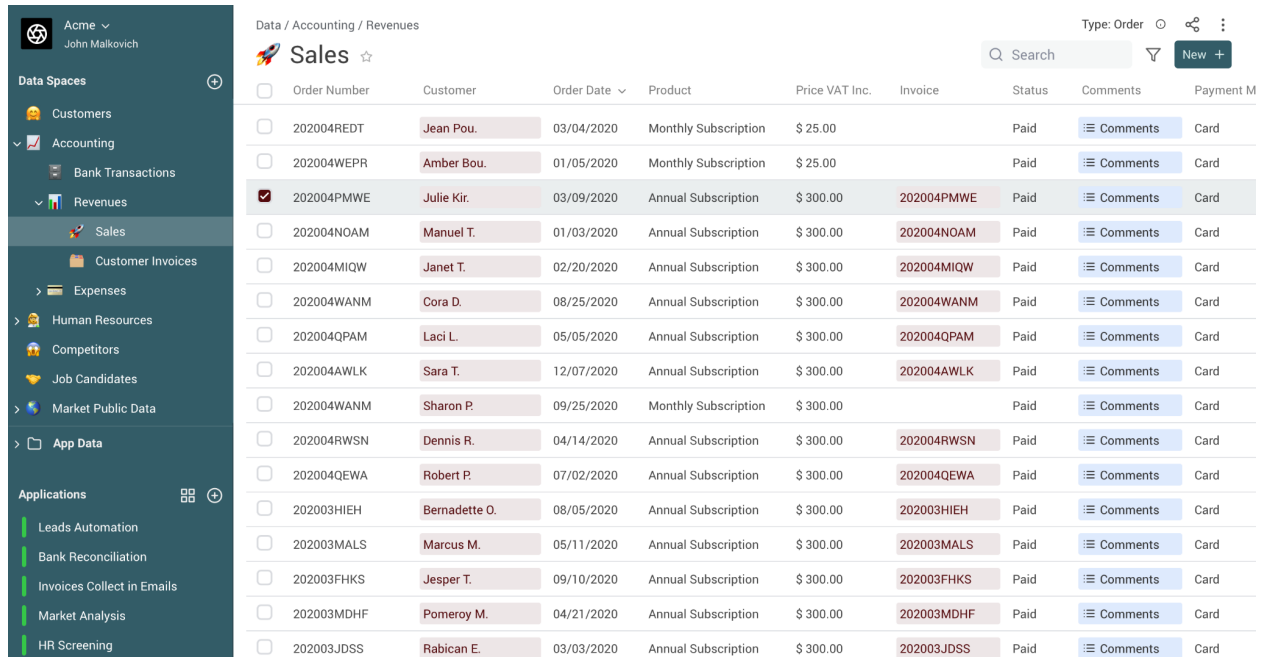
from the imported file exists in the collection or elsewhere in the file, the data will be merged or deleted according to the merging mode.

Pros	Cons
<ul style="list-style-type: none">- user-friendly- free- lots of customization options	<ul style="list-style-type: none">- external app (no Salesforce integration)- must explicitly specify the editing interface for editing each item- older interface

Datablist offers a free version and a monthly subscription version at \$12 per month per user. The primary differences between the free and paid version are explained in the table below

Free Version	Paid Version (Includes everything from the free plan and:)
Unlimited Collections	Import up to 1.5 million items per collection stored in your browser storage
Import up to 1 million items per collection stored in your browser storage	Sync collections with less than 100,000 items in the cloud
Sync collections with less than 1,000 items in the cloud	Unlimited Team Members
Duplicates Finder	Access to all data enrichments
Access to free data enrichments	Priority support
Import CSV and Excel files	
Export to CSV and Excel files	

Maryland Hillel Data Architecture Review



The screenshot displays the Insycle CRM interface. On the left is a dark sidebar with a navigation menu. The top of the sidebar shows the company name 'Acme' and the user 'John Malkovich'. Below this, the 'Data Spaces' section includes 'Customers', 'Accounting' (selected), 'Bank Transactions', and 'Revenues'. Under 'Revenues', 'Sales' is selected. Other options include 'Customer Invoices', 'Expenses', 'Human Resources', 'Competitors', 'Job Candidates', 'Market Public Data', and 'App Data'. The 'Applications' section at the bottom lists 'Leads Automation', 'Bank Reconciliation', 'Invoices Collect in Emails', 'Market Analysis', and 'HR Screening'. The main content area is titled 'Data / Accounting / Revenues' and 'Sales'. It features a search bar, a 'Type: Order' dropdown, and a 'New +' button. Below this is a table with columns: Order Number, Customer, Order Date, Product, Price VAT Inc., Invoice, Status, Comments, and Payment M. The table contains 15 rows of data, each representing a sales order. The third row is highlighted with a red background.

<input type="checkbox"/>	Order Number	Customer	Order Date	Product	Price VAT Inc.	Invoice	Status	Comments	Payment M
<input type="checkbox"/>	202004REDT	Jean Pou.	03/04/2020	Monthly Subscription	\$ 25.00		Paid	Comments	Card
<input type="checkbox"/>	202004WEPR	Amber Bou.	01/05/2020	Monthly Subscription	\$ 25.00		Paid	Comments	Card
<input checked="" type="checkbox"/>	202004PMWE	Julie Kir.	03/09/2020	Annual Subscription	\$ 300.00	202004PMWE	Paid	Comments	Card
<input type="checkbox"/>	202004NOAM	Manuel T.	01/03/2020	Annual Subscription	\$ 300.00	202004NOAM	Paid	Comments	Card
<input type="checkbox"/>	202004MIQW	Janet T.	02/20/2020	Annual Subscription	\$ 300.00	202004MIQW	Paid	Comments	Card
<input type="checkbox"/>	202004WANM	Cora D.	08/25/2020	Annual Subscription	\$ 300.00	202004WANM	Paid	Comments	Card
<input type="checkbox"/>	202004QPAM	Laci L.	05/05/2020	Annual Subscription	\$ 300.00	202004QPAM	Paid	Comments	Card
<input type="checkbox"/>	202004AWLK	Sara T.	12/07/2020	Annual Subscription	\$ 300.00	202004AWLK	Paid	Comments	Card
<input type="checkbox"/>	202004WANM	Sharon P.	09/25/2020	Monthly Subscription	\$ 300.00		Paid	Comments	Card
<input type="checkbox"/>	202004RWSN	Dennis R.	04/14/2020	Annual Subscription	\$ 300.00	202004RWSN	Paid	Comments	Card
<input type="checkbox"/>	202004QEWA	Robert P.	07/02/2020	Annual Subscription	\$ 300.00	202004QEWA	Paid	Comments	Card
<input type="checkbox"/>	202003HIEH	Bernadette O.	08/05/2020	Annual Subscription	\$ 300.00	202003HIEH	Paid	Comments	Card
<input type="checkbox"/>	202003MALS	Marcus M.	05/11/2020	Annual Subscription	\$ 300.00	202003MALS	Paid	Comments	Card
<input type="checkbox"/>	202003FHKS	Jesper T.	09/10/2020	Annual Subscription	\$ 300.00	202003FHKS	Paid	Comments	Card
<input type="checkbox"/>	202003MDHF	Pomeroy M.	04/21/2020	Annual Subscription	\$ 300.00	202003MDHF	Paid	Comments	Card
<input type="checkbox"/>	202003JDSS	Rabican E.	03/03/2020	Annual Subscription	\$ 300.00	202003JDSS	Paid	Comments	Card

Insycle (<https://www.insycle.com/why-insycle/>)

Insycle provides a comprehensive customer data management solution, streamlining the process for businesses to oversee, automate, and uphold well-organized customer databases. This empowers their teams to work more effectively, enhancing reporting accuracy and ensuring data integrity and proper formatting.

1. **Data Cleansing and Enrichment:** Insycle offers tools for data cleansing, deduplication, and enrichment. Users can identify and merge duplicate records, standardize data formats, and enrich data with additional information from external sources.
2. **Bulk Data Editing:** The platform allows for bulk editing and updating of CRM data. Users can perform mass updates, edits, and deletions across multiple records simultaneously, saving time and effort.
3. **Data Standardization:** Insycle provides tools to standardize data by enforcing consistent formats, values, and naming conventions across various fields. This helps maintain data consistency and accuracy.
4. **Advanced Filtering and Segmentation:** Users can create complex filters and segments to isolate specific subsets of data for analysis or management. This feature enables targeted actions on specific groups of records.
5. **Workflow Automation:** Insycle offers workflow automation capabilities, allowing users to create and automate workflows for data management tasks. This can include scheduled tasks, rule-based actions, and triggers based on specific conditions.
6. **Data Import and Export:** Users can easily import data into their CRM platform from external sources or export CRM data for analysis or backup purposes. Insycle supports various file formats for data import/export.
7. **Data Validation and Error Identification:** The platform helps identify data quality issues by flagging errors, inconsistencies, and missing information within the CRM data. This assists in maintaining data accuracy.

Maryland Hillel Data Architecture Review

Pros	Cons
<ul style="list-style-type: none">- Fairly cost-efficient (\$1-2 per 1,000 records per month)- automatic deduplication capability for Salesforce- merge contacts with existing records or generate new accounts no present using any specific identifier that is required- advanced matching rules to allow duplicates to be compared by various properties	<ul style="list-style-type: none">- UI may be difficult to navigate- you have to pay separately for different platforms you use within Insyncle- terminology for filtering may differ from Salesforce

Cohesity (<https://www.cohesity.com/solutions/>)

Cohesity is a data management platform that offers a range of features aimed at simplifying and consolidating data management, protection, and utilization.

1. **Data Protection and Backup:** Cohesity provides comprehensive backup and recovery solutions for various data sources, including physical servers, virtual machines, databases, and cloud-native applications. It offers efficient snapshots, incremental backups, and rapid recovery options.
2. **Global Deduplication and Compression:** Cohesity uses advanced data reduction techniques such as global deduplication and compression to optimize storage efficiency, reducing the overall storage footprint.
3. **Analytics and Insights:** Cohesity provides analytics and reporting capabilities, offering insights into data usage, storage patterns, and compliance adherence. This helps optimize storage resources and make informed decisions about data management strategies.

To pay for the system, Cohesity costs \$1,200 for the whole year. Their server costs \$90,000 however this should not be necessary for the purpose of this data management system.