Paxata™

Welcome to Paxata!

Document version 2.13

Date: August 01, 2016

Paxata is a business analyst solution to find, combine, enrich, and shape the right datasets for analytical needs—which can be consumed in your visual BI tools of choice.

Table of Contents

Paxata Adaptive Data Preparation™

Empowers every business analyst to **connect, explore, transform and combine** data on their own or work with peers in a shared, transparent environment

www.paxata.com

What is Paxata?

Paxata is the first Adaptive Data Preparation™ platform built for the business analyst. Our technology dramatically reduces the most painful and manual step of any analytic exercise—turning raw data into ready data for analytics—and empowers analysts to drive greater value for the business. Paxata leverages the power of sophisticated software to make the production of rational, relevant datasets a simple task for business analysts that takes minutes, not months. Paxata allows people like you to tackle complex preparations of data without the need for IT support or specialized skill sets.

Who is an Analyst?

An analyst is anyone who needs to examine a set of data to learn something new. Analysts produce answers or, at the very minimum, they produce data products that are used by others to find answers.

There is no single type of analyst; analysts have many job functions and titles. However, analysts usually have at least one of the following functions associated with their work duties:

- Production: This function may deliver answers to others or it may deliver sets of data that others use to make decisions. The production process may deal with a single system or it may sit at the intersection of multiple systems. The production role often requires repetitive tasks every time the analyst encounters a new set of data. An analyst in this role may also use Tableau™ or QlikView® as part of the production process.

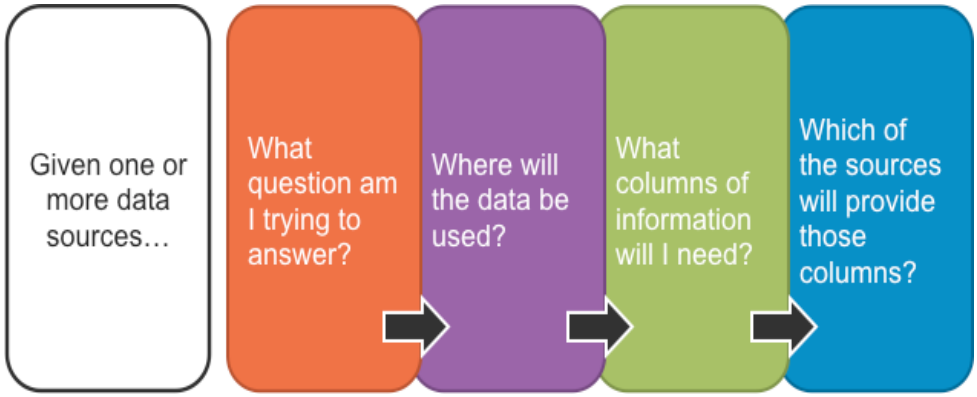
- Inspection: This is the gatekeeper function. In this role, the analyst may be the data entry point for the organization—"We have a new list of leads from a marketing company!" Or, the analyst may curate the information in an existing data store such as a database or Apache™ Hadoop®. In all cases, an analyst with an inspection role needs to determine whether or not the company's data is in a useful form, and whether or not the data adds value to the information already in use.

- Collaboration: Rather than working alone, an analyst with this function needs to manage their work in coordination with others. This may require the analyst to pass work-in-progress onto others for additional data preparation steps. Or this may require the analyst to request a cohort's assistance when working through a specific data preparation exercise.

How Do Analysts Use Paxata?

No matter what function you perform as an analyst, Paxata helps you to explore your data, remove irregularities, form it into new shapes, and combine it with other datasets in order to efficiently prepare meaningful data for your particular needs. Paxata does not limit you to pre-defined workflows. If you know the outcome you're looking for or have a particular question you need to answer, those objectives will drive the ways in which you use Paxata to transform your data.

The steps outlined below explain how your objectives can drive your data preparation steps in Paxata.



1. What question am I trying to answer? A user must either start with a goal or, after exploring the data in Paxata, be able to develop one. Almost any data preparation-related need will qualify:

   - *A retail business process improvement initiative needs to correlate gross margin and volume by distributor. The source systems don't match up—and they are too big to use in Excel™.*
   - *Marketing wants to measure the incentive programs on distributor sales. This involves tying the general ledger of the sales organization to external sales data, and then comparing year over year results.*
   - *The sales organization wants better insight into the products that customers are buying by Business Unit, and help determining which products might be presented for promotions.*

2. Where will the data be used? The intended use of the prepared data helps to define the steps that a user must accomplish in Paxata. For example, data to be visualized as a pie chart in Tableau™ will likely need to have different characteristics than information presented in a summarized table.

   - *Names of customers in a CRM do not match up with customers in the invoicing and billing system. A consolidated list of difference needs to be produced so it can be investigated and corrected.*
   - *The summary information that a team receives for monthly sales by category cannot be used to generate a bar chart displaying trends over several years. The monthly sales data needs to be combined and put into a form where a visualization tool can easily display the data graphically.*
   - *A large list of global products needs to be broken down into smaller tables and organized by availability within a geographic region. The regional managers will use the information as checklists to verify they have appropriate product mix.*

3. What columns of information do I need? Some people refer to this step as determining the *shape* of the data. It involves not only determining which columns should be included but also how the information in the columns should be structured.

   - *In order to understand the cost of hiring, information about job applicants, the positions for which they accepted offers, and the fees charged by recruiters must all be part of the final data set.*
   - *An examination of the sales performance for individuals in various districts must include individual names, sales figures, and the districts to which each belongs.*
   - *Preparing to mail out special offers involves not only names and addresses but also information about the types of offers recipients are supposed to receive.*

4. Which data sources provide the necessary columns of data? Sometimes a single set of data will have all of the necessary information for analysis. Often, this information must be obtained from separate systems or people.

   - *The billing system provides customer IDs on invoices, but the sales representative for each customer is missing and must be imported from the company's custom sales tracking tool.*
   - *An inventory list contains part numbers, but the complete list of nomenclatures used to refer to each part can only be found in the company's ERP application.*
   - *The software products licensed by a customer exist in the CRM platform, but the number of incidents they opened with the Support team can only be found in the IT service management database.*

Resources Available to You

This Welcome Document is intended for new users who need a quick familiarization with Paxata that highlights features and functions within the application.

- Overview via web meeting
- On-site Training
- Quick Reference Guides
- Paxata Support
    - Submit support tickets at servicedesk.paxata.com
    - Email support directly at servicedesk@paxata.com

Getting Started

Note: Technical documentation and support materials include details based on the full set of capabilities and features of a specific release. Please note that individual access to specific functionality may vary based on deployment and license types.

Paxata is designed to be simple to learn and use. However, because Adaptive Data Preparation is a new concept, it may be helpful to have a quick overview of how to get started. This document highlights the major areas of functionality, and demonstrates some of the more commonly used features.

Figure 1 | *Login with the credentials provided by your company's Paxata administrator. Contact the Service Desk if you need help.*

To begin, login with the credentials provided in your "Quick Reference Sheet" (page 2). If your username or password is not there, check with the Paxata administrator at your company, or send an email to the Paxata Service Desk (servicedesk@paxata.com).

Note: Paxata currently supports the following browser versions for both Mac and Windows:

Mozilla Firefox: Extended Support Release (ESR) 38.6.1 for Mac and Windows
Google Chrome: 48 for Mac and Windows
The recommended resolution for the Paxata application is 1024x768

Once logged in, the “Home” screen, displays. You’ll want to take special note of three “landmarks” you can count on seeing from anywhere within the application:

- **Application Menu**: click to access the Data Library, Paxata Projects and any other menus that your Paxata user rights allow you to access.
- **User Menu**: click to display the options for changing your password and logging out
- **Help Panel**: click to reveal and hide the help panel

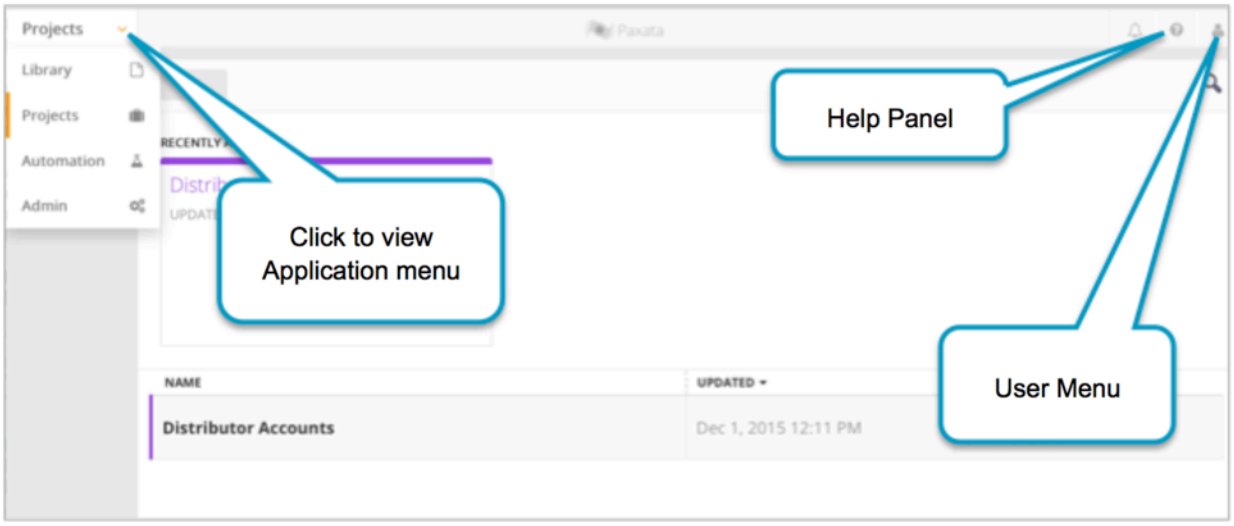


Figure 2 | *Take note of three "landmarks" available to you from anywhere in the Paxata application.*

1. The Basics

Before beginning data preparation, let's begin with some basics. You'll want to know how to: change your password and recognize common icons.

(i.) Your password

It's a good idea to login and change your password as soon as you get your Paxata account. To change your password, select the "my account" option in the User Menu.

Login with your credentials provided in the "Quick Reference Sheet" (page 2). To change your password: click the user menu icon in the top, right corner of the screen. Then click on "my account" from the menu that appears. Enter your new password and confirm it. Then click the "Save" button. (Note that only administrators can change user roles.)

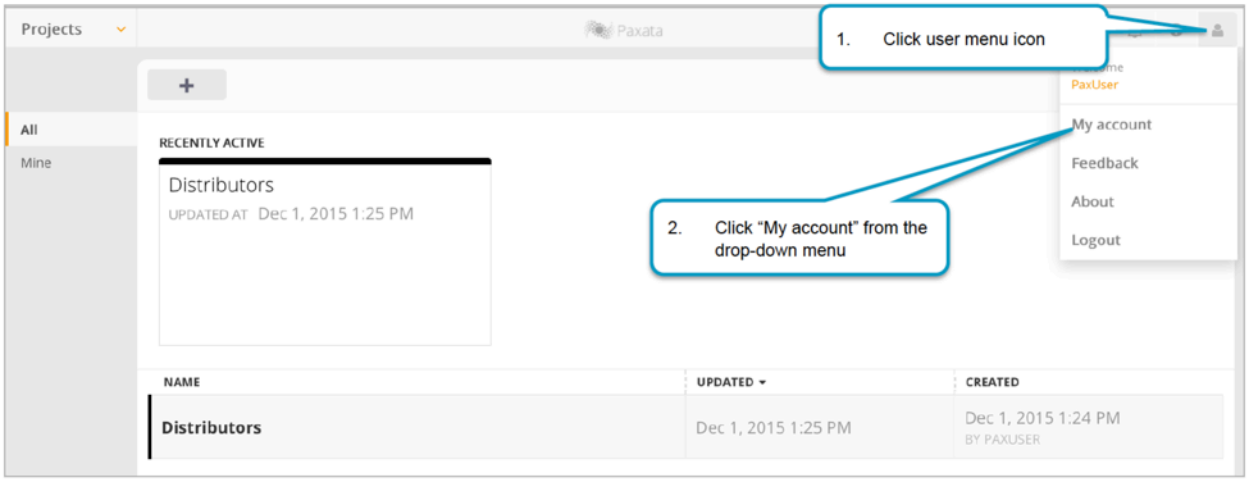


Figure 3 | *It is recommended that changing your password is your very first step upon initial login to Paxata; this will help keep your data and Projects more secure.*

(ii.) Common icons

The following table helps you to recognize and identify the functions associated with icons seen throughout the Paxata application.

Icon	Meaning
	Any time you see a plus icon, you'll click it to create; for example a new Project or data set
	Clicking the magnifying glass allows you to search
	The eye icon toggles visibility allowing you to view or hide something
	A pencil icon allows a value to be viewed and changed
	When an "x" is shown, it means something can be deleted, removed, or closed
	Indicates either text or text plus other data types in the data set column
	Column with values composed entirely or primarily of numbers
	A column with this icon indicates it stores dates and time values

2. Understanding the Data Library

The Data Library is where Paxata stores all datasets and AnswerSets. It's the repository from where you'll pull base data (datasets) for preparing your Projects and the location where your prepared AnswerSets are saved. From the **Application** menu, open the Library.

When the Data Library is open, datasets can be imported into it by [A] clicking the "plus" icon. Depending on the configuration of your particular system, several import options are available. Existing datasets can be searched with options [B] available in the "Show Only" panel. Moving your mouse over a dataset in the Library [C] highlights it and allows for actions to be taken directly on it, including [D] triggering export actions.

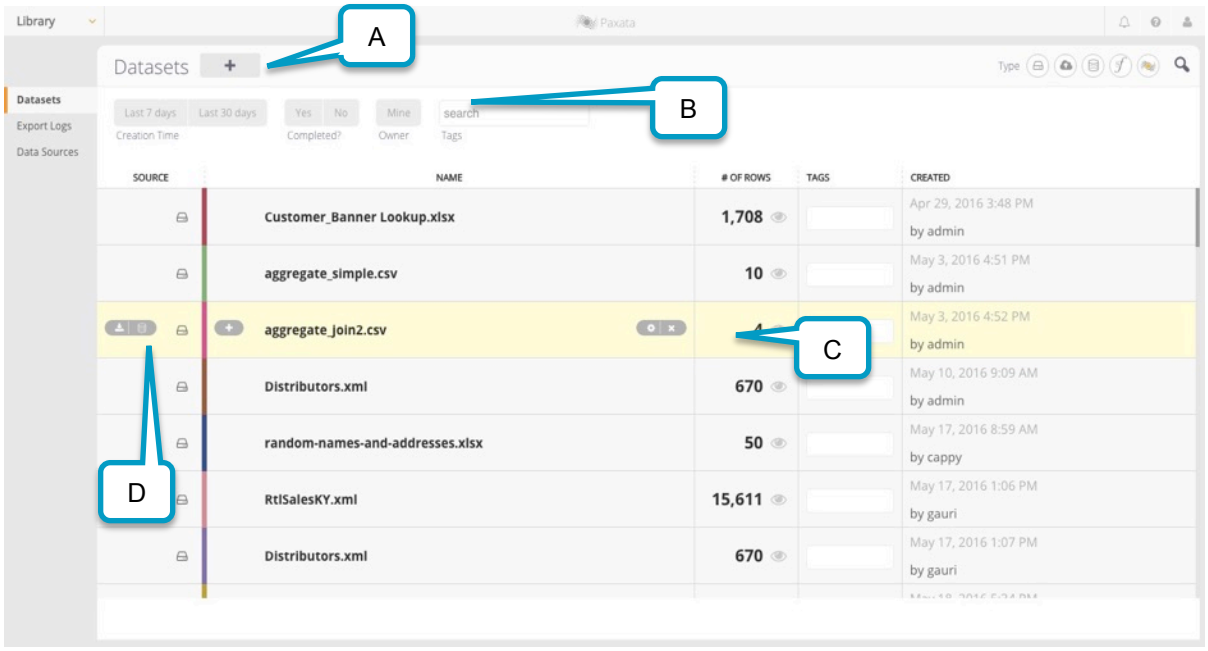


Figure 4 | *The Data Library is the dataset repository for all users of Paxata. It provides mechanism for import and export and several mechanisms for searching.*

(i.) Importing

Paxata's connector framework allows you to import from a multitude of data sources and file types.

When you click the “plus” icon (as seen in *Figure 4*), you are brought to the following screen where you select the data source that you want to import data from. You can then browse to the particular dataset(s) that you want to import into Paxata’s Data Library.

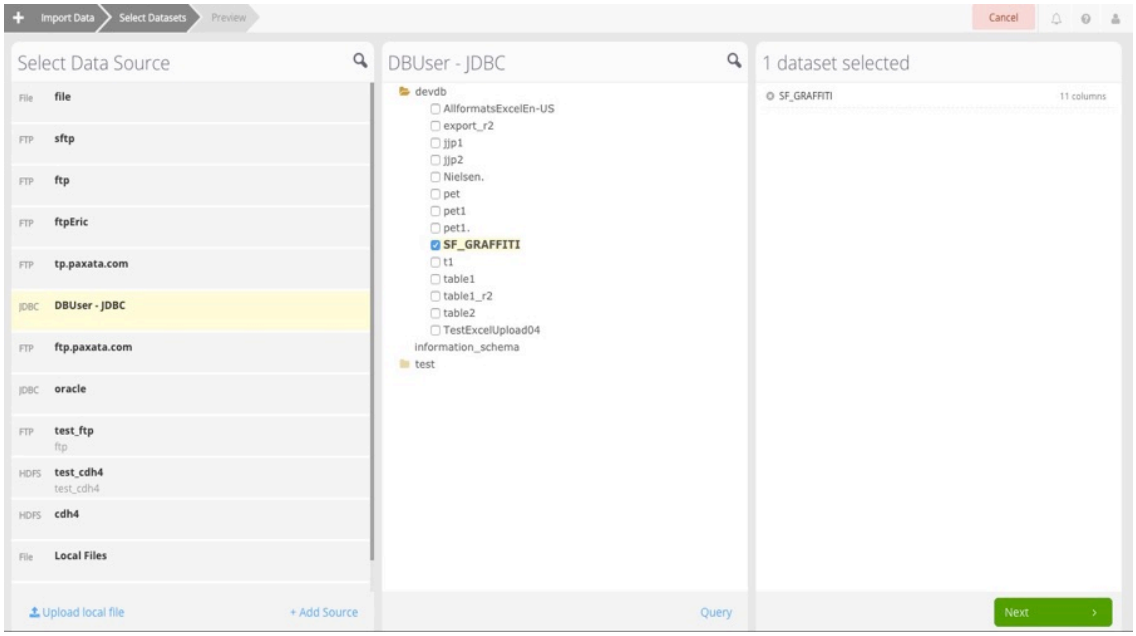


Figure 5 | *Select your data source and navigate directly to the dataset(s) you want to import.*

After you navigate to your desired dataset(s) for import and select your parsing options, click “finish” to complete the import process.

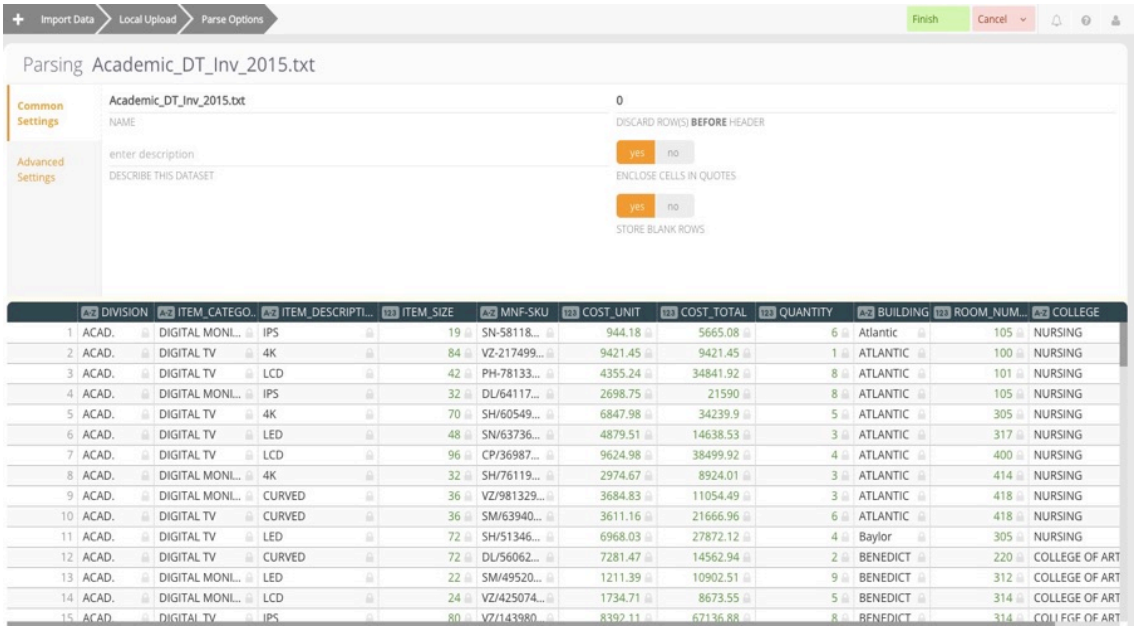


Figure 6 | *After loading the data, but before it is available for use in the Data Library, you must select import options (or accept the defaults); the choices here control how Paxata parses a particular data source.*

(ii.) Exporting

Paxata's connector framework allows you to export your AnswerSets from the Data Library to a multitude of data sources and formats.

When exporting an AnswerSet as a file, the Export Settings can be configured as shown below. After your selections are complete, you are automatically taken to the Data Library Export Log in order to view the progress of your export(s).

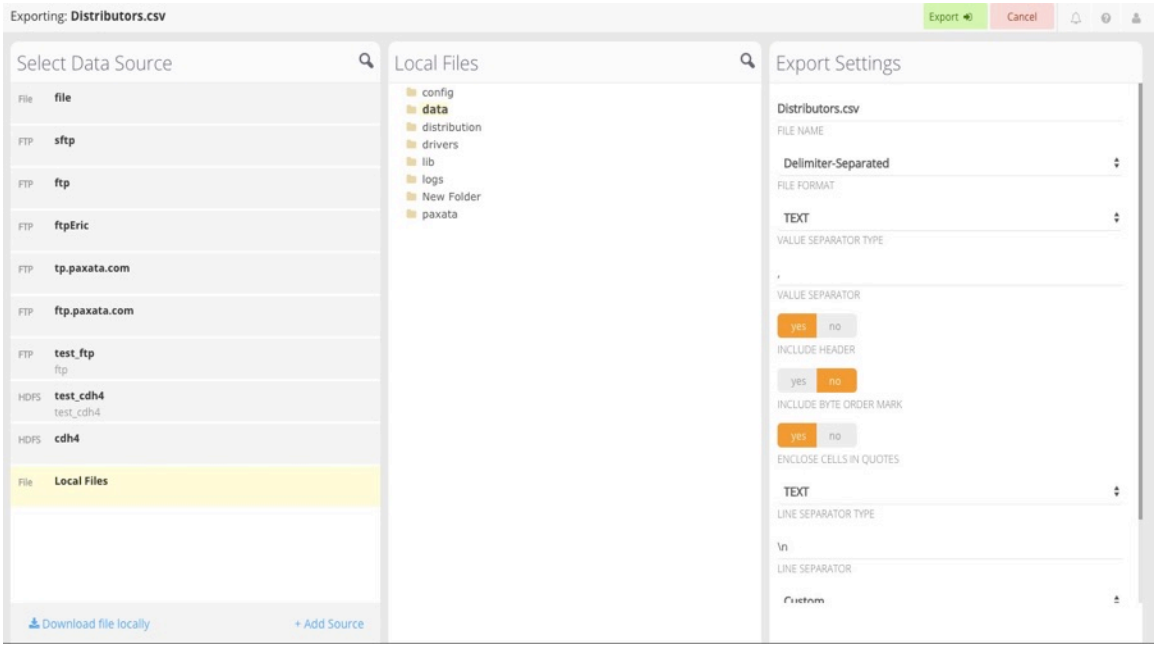


Figure 7| *Whenever the green "Export" button is clicked, the export is triggered and you are automatically taken to the Export Log in order to view the status of your export(s).*

3. Understanding Projects

All data preparation activities take place inside of a Project. The Project inventory screen is the first screen after you login. It can also be reached by clicking Projects from the **Application** menu.

To open an existing Project, click on the box with the name of a recent Project [A] at the top of the screen, or [B] search through the lower panel and click on the desired Project in the list. Project names and descriptions can be edited from here, and Projects can also be deleted. New Projects are started from the inventory screen by clicking the "plus" icon [C] adjacent to the word "Projects".

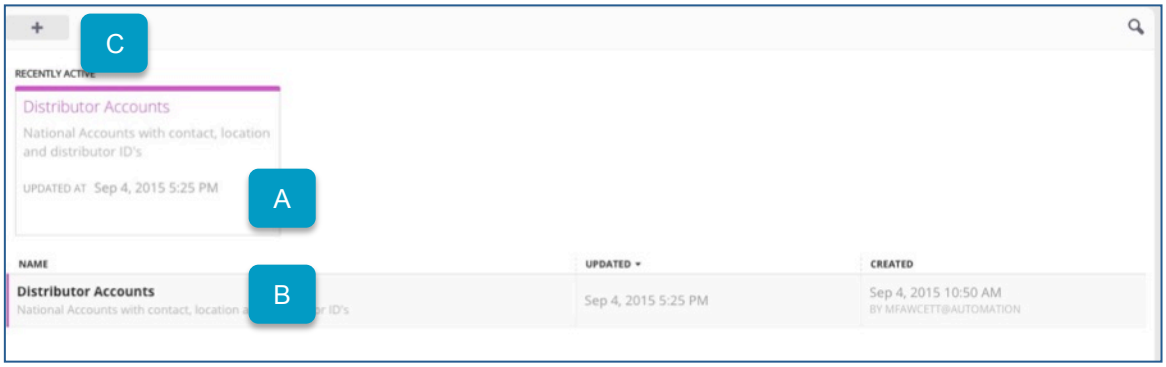


Figure 8 | *The Project inventory screen is populated as projects are created. Existing projects can be accessed via this screen or new projects can be created.*

(i.) First steps

When creating a new Project, it must be given a name; the description is optional. Clicking the "Save" button creates the Project and returns you to the Project inventory screen. Selecting "Save and Open" creates the Project and launches it for you to begin working.

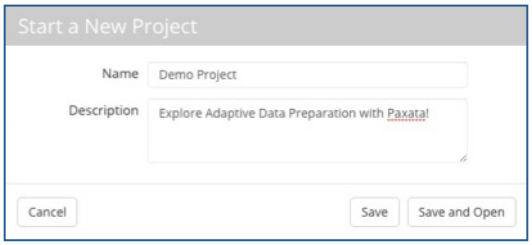


Figure 9 | *The Start a New Project dialog box begins your Adaptive Data Preparation™ journey!*

The first task that you must complete after creating a Project is adding the "base". The base data forms the foundation of the Project and is the data against which all other actions in the Project will be performed.

Note that you can choose an entire dataset when you import the base dataset, or you can choose to Sample a dataset from your Paxata Data Library. Sampling is useful if your IT group has set an import size limit for Paxata Projects and you want to work with a dataset that exceeds the limit. For an optimal sample, your dataset should exceed 100k rows.

When you [A] click the "click to select" button in the Project, you are taken to the Data Library. (If the desired set of data is not in the Library, you can import as explained in section 2i above.) Locate your base dataset in the Library, then [B] click the green "Select" button that appears just to the left of the dataset's name when your mouse hovers over the row.

Figure 10 | *The Base dataset forms the project foundation; all activities in the project impact this dataset. It is added from the Data Library as the very first step in beginning a data preparation project.*

Once the base dataset is added to the Project, it's initially in preview mode. This allows you to verify the dataset is the appropriate one for your base set purposes. After you examine the dataset, you MUST remember to click the "Save" button in the upper right-hand corner of the screen to commit the dataset to the Project and begin your work.

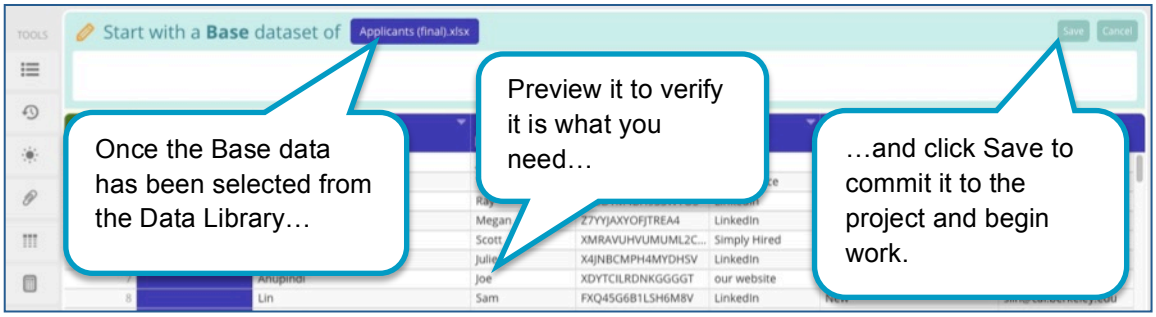


Figure 11 | REMEMBER: *The dataset is not fully loaded into the project until after the "Save" button is clicked. Paxata is simply previewing the data. Once you select it, you must save the base dataset as part of the project before beginning preparation activities.*

(ii.) Column operations

Within a Project, column operations are activities that are triggered by [A] hovering your mouse over the drop-down arrow in a particular column and [B] selecting an option from the menu that appears.

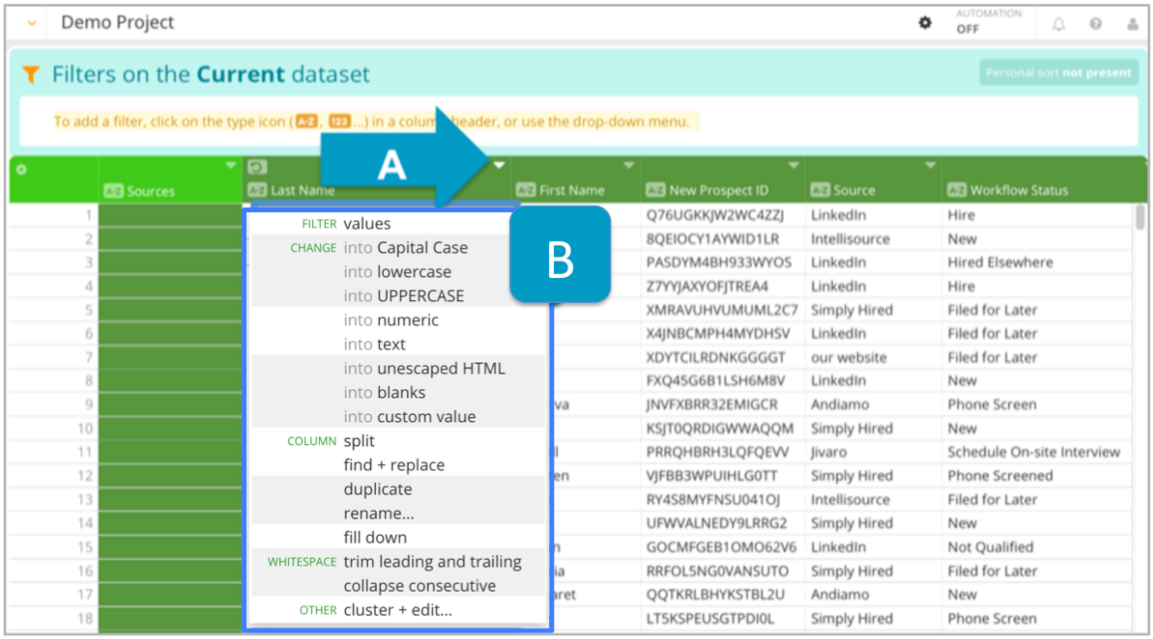


Figure 12 | *Column operations are available as options within the menu displayed from the drop-down arrow.*

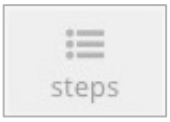
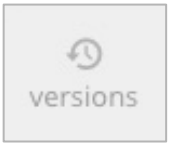
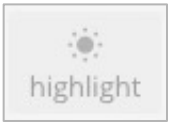
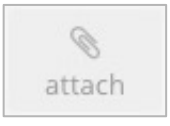
A column operation impacts only the column to which it is applied. The table below summarizes the tools available within this menu.

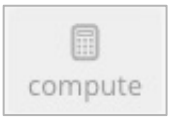
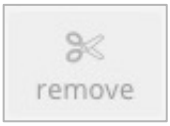
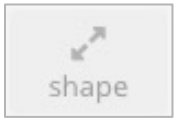
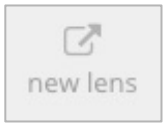
Group	Operation	Description
FILTER	by values	Launches the Filtergram panel for this column at the top of the screen. The Filtergram panel displays the distribution of values in this column as a histogram and provides tools for dynamically filtering your data.
CHANGE	into Capital Case	The first character of every word in each cell is capitalized
	into lowercase	All text becomes lowercase
	into UPPERCASE	All text becomes uppercase
	into numeric	For numbers stored as text, converts them into numeric values suitable for mathematical calculations
	into text	For numeric values, converts them into text characters suitable for use in string operations
	into unescaped HTML	Transforms all HTML character codes into the actual characters they represent
	into blanks	Empties all values, effectively creating a column of blank cells
	into custom value	Selected cells are changed into a user specified custom value

Group	Operation	Description
COLUMN	split	Separates a single column into multiple columns
	find + replace	Substitutes one set of characters for another
	duplicate	Creates a copy of the column
	rename…	Allows the column heading to be changed
	fill down	Use fill down to populate blank cells in this column with data from the cell above. The fill down continues for blank cells until the next populated cell is encountered in the column. As part of this operation, you also have the option to bind this column to the sort order for other, existing columns in your dataset. This is useful when you need to confirm a sort order for your dataset before further exploration. See "Auto Number" in the next section below for more details on assigning sort order to columns.
WHITESPACE	trim leading and trailing	Removes spaces from the beginning and end of the cell
	collapse consecutive	Takes all inter-word spaces and consolidates them to a single space
OTHER	cluster + edit…	Takes groups of similar text values and converts them all into identical pieces of text

(iii.) Dataset operations

Along the left-hand side of the Project are a number of buttons that contain operations that impact the entire dataset. Some of these operations have groups of functions within them.

Button	Operation	Sub-Function	Description
	Display Steps		Opens the Steps panel, allowing for "undo" activities; also supports modifying and reordering preparation steps
	Display Versions		Opens the Versions panel, allowing you to navigate between different versions of your Project
	Highlight	• Highlight patterns • Highlight spaces • Highlight ranges	Colors visible cells to assist in identifying data within the grid that might otherwise be difficult to discern
	Attach	• Lookup • Append	Combines columns from additional datasets with the columns already existing in the Project (*aka*, "Combine" or "Join") or appends additional data to already existing columns

Button	Operation	Sub-Function	Description
	Columns		Allows for columns to be reordered and removed
	Create Computed Column		Creates a new column using programmatic formulas; may use data from existing columns to produce the result
	Remove Rows		Removes rows of data based on a matching pattern of data within those rows
	Sampling		Allows you to add a data prep step to sample your dataset. The sampling tool gives you the flexibility to filter down to a specific set of rows in your data, and then sample on the remainder.
	Shape Dataset	• Deduplicate • Group by • Transpose • Pivot • Depivot	Advanced features that allow you to reshape your data, including: aggregations, deduplicating records, and pivot tables.
	Auto Number		Provides a unique index number for every row in your dataset. You can then bind the column to the sort order for other, existing columns in your dataset. Auto Number is useful when you need to: • track your dataset's original order • assign row ID's to your dataset
	New Lens		Inserts a new Publishing Lens in your Project

4. Key Functionality

There are several areas within Paxata that provide tremendous value for the data preparation process. The sections below describe where to find these features and how to use them.

(i.) Lookups

A Lookup combines columns from additional datasets with the columns already existing in the base dataset. You may know this feature as "Combine" (because the operation combines two different datasets) or "Join" (from a similar SQL operation). If you are a Microsoft Excel® user, you may be familiar with this as a VLookup—but in Paxata, it's without the restrictions that VLookup normally imposes.

When the Lookup function is triggered, the first step is to [A] select a dataset from [B] the Data Library to bring into the Project. This is performed with the same steps you followed to bring in your base dataset.

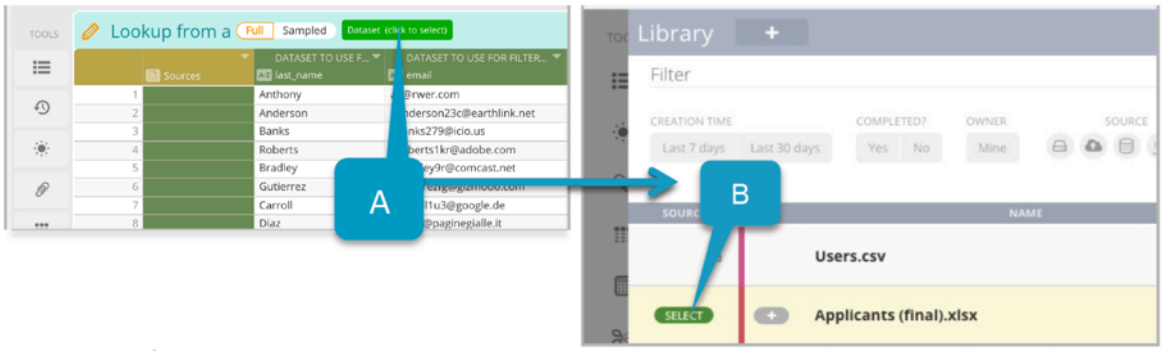


Figure 13| *A Lookup dataset is selected the same way the Project's Base dataset was chosen; the Data Library is displayed and the green "Select" button next to the desired dataset is clicked.*

After the Lookup dataset is selected, you must choose the columns on which you wish to combine. These should be columns that have something in common. In other words, the data should have a shared meaning and form so that it serves to link the lookup data to the base data already in the Project. For example:

- A dataset with banking information and a list of bank customers might be combined on the account number; this allows the joined dataset to show which customers own specific bank accounts.
- A job applicant's first name, last name, and address may be used to uniquely identify them when information from the job posting systems is linked with records used for the hiring process.

After selecting a Lookup dataset, Paxata examines the columns in both datasets to [A] suggest possible columns on which to combine, and it displays them [B] based on the percentage of overlap between the two columns. You can accept the Paxata suggestion, or you can select your own columns on which to join. Use the [C] Options to change the Lookup type that is being conducted, the [D] preview panel to verify the correct lookup, and then [E] click the "Save" button to commit the Lookup.

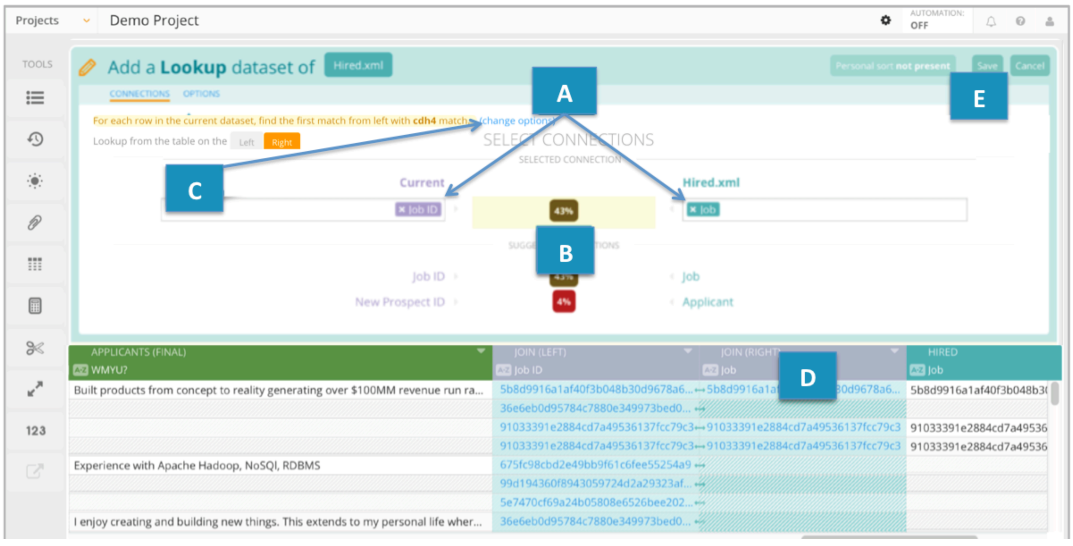


Figure 14| *Paxata examines the columns in both your base and lookup datasets and suggests possible columns on which to combine. You can accept the suggestion or select your own columns on which to join.*

(ii.) Append

Append allows additional data to be added (appended) to already existing columns within the base data. When Append is launched, the first step is to [A] select a dataset from the [B] Data Library to append to your base data. This is performed in the same way that your base data was brought into the Project.

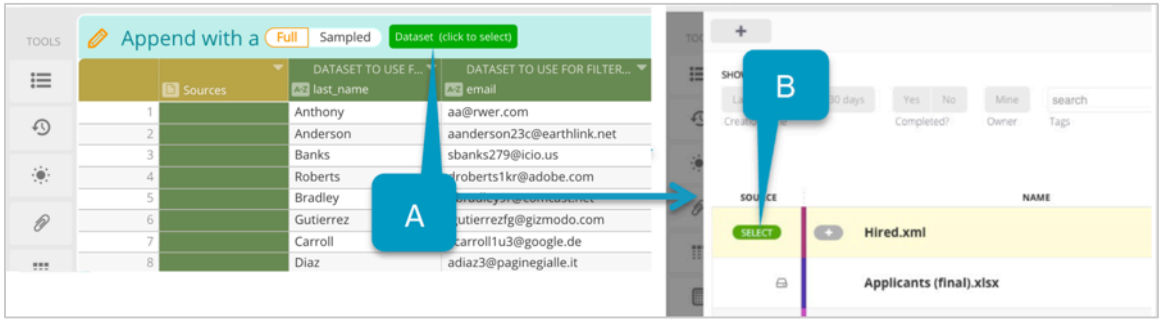


Figure 15| *To conduct an Append, a dataset is selected in the same way you add a Base dataset or Lookup Table.*

Once a dataset has been selected, Paxata automatically matches columns [A] with the same names. If any of these matches need to be undone, simply click the X button [B] between the matches. You can manually match columns [C] together as well via the drop downs. Once you are satisfied with the matches, click Save [D] to commit the appended data to the Project.

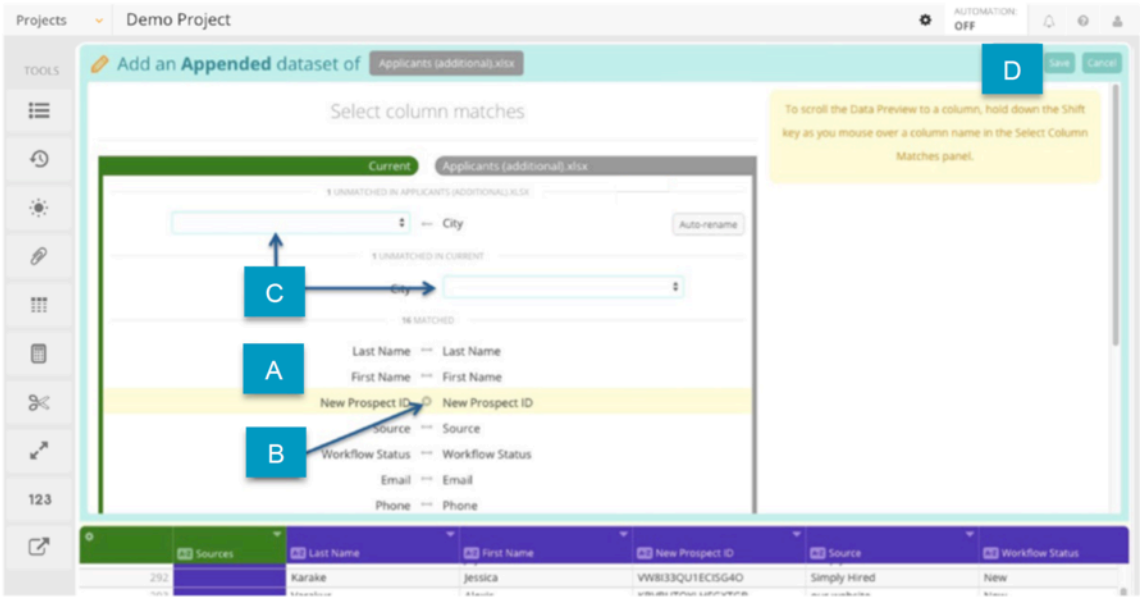


Figure 16| *Paxata will automatically match columns with the same name during an Append, but you can manually remove these by clicking the "X" between each match, and you can also specify your own matches via the drop down menus. Don't forget to click "Save" to commit this action to the project.*

(iii.) Steps

The Steps panel allows for an extensive range of options regarding how your data is transformed. For most beginning users, however, it serves two (2) primary purposes:

1. Review all data preparation steps within the Project

2. Delete unwanted or inadvertent data preparation steps from the Project

When the Steps button is clicked, the Steps panel pops-out from the left-hand side of the screen. By default, it activates in read-only mode and shows all steps taken in the Project. This allows you to review of all Project steps already defined.

In order to make changes to existing steps and/or delete a step previously completed, the Steps panel [A] must be placed into “Edit” mode. Once in edit mode, mouse-over the step to be removed and [B] click the “x” on its left-hand side to remove it (prior to entering “Edit” mode, Paxata will display an “eye” icon rather than an “x”, this will allow you to mute the step without deleting it). Finally, [C] click the “Save” button to commit the changes.

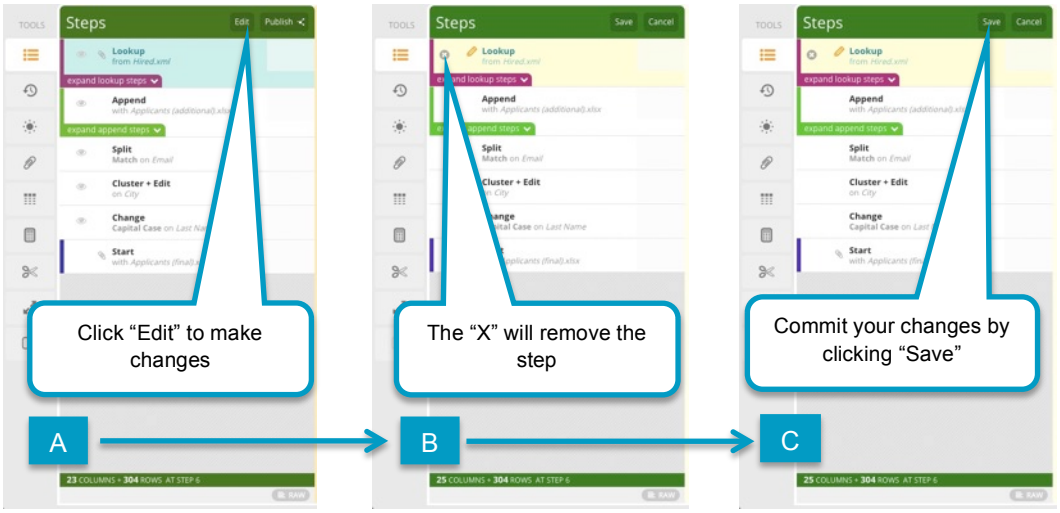


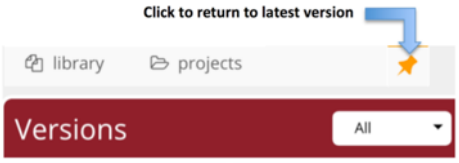
Figure 17 | *The Steps panel is a powerful feature that allows granular changes to be made to the data preparation process. This image shows the Steps panel in three successive modes: in "Review" mode [A]; in "Edit" mode [B]; and with changes made that need to be saved [C].*

(iv.) Versions

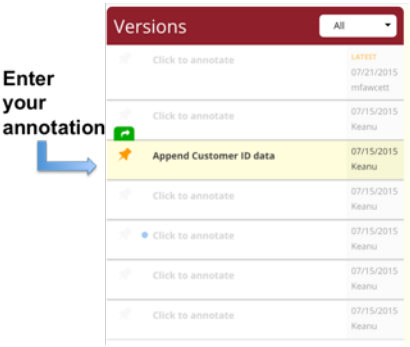
Every time an action is taken in your Project—for example adding a step, removing a step, re-arranging steps—a new version of your Project is created. All versions are listed in the **Versions** panel. Click any version in the panel to view your Project at that point in its history.

The latest version of your Project is listed at the top of the panel. A blue dot appears adjacent to any version that has been published as an AnswerSet to the Data Library:

Any time you are viewing a previous version of your Project, an orange pin icon is displayed at the top of the Versions panel to remind you that you're not viewing the most recent version. If you click the pin icon, you are immediately returned to the most recent version of your Project.



You may find it helpful to annotate a Project version to reflect the steps you took to render that particular version, for example: "Append Customer ID data." To annotate a version, select the "Click to annotate" text for the version and provide an annotation:



(v.) Publishing and Lenses

Data within a Project is not available for use until it is published. The publishing activity produces a copy of the Project's data, with all transformations applied, as a new dataset within the Data Library. From here, the newly transformed dataset can be used in other Projects or exported for use outside of Paxata.

Lenses are used to create publishing points from Steps in your Project. When you publish from a Lens, the resulting AnswerSet is a snapshot of your dataset at that particular Step in the Project. By default, the AnswerSet is saved to your Data Library.

A Lens can be added to any Step or sub-Step in your Project, for example, to an Import Step of an Append. An existing Lens can be dragged to any Step. Additionally, multiple lenses can be applied. The Lenses you create are retained in Project versions, and you can publish AnswerSets from Lenses in older versions of your Project.

Note that all Lenses are public to anyone who shares your Project.

Lenses are also essential for Project Automation because they define the publishing points to use for automated jobs. When you set up a Paxata Project for automation, you select Lenses and configure a corresponding schedule to automatically publish AnswerSets to your Data Library. Therefore, in order to automate a Project, you require at least one Lens in the Project.

(vi.) Data Library and Project Automation

There are two types of workload automation that Paxata provides to reduce the number of repetitive tasks taken to produce AnswerSets: Library automation and Project automation.

- Library automation - This type of automation enables an existing Data Library file to automatically pull an update from its original source based on the schedule you configure. Data Library files from the following sources can be automated: JDBC, HDFS and Salesforce.

- Project automation - This type of automation allows you to schedule a Project to automatically publish an AnswerSet in the Data Library based on the schedule you configure. The AnswerSet can also be exported to Hive or HDFS as part of the automated publishing process.

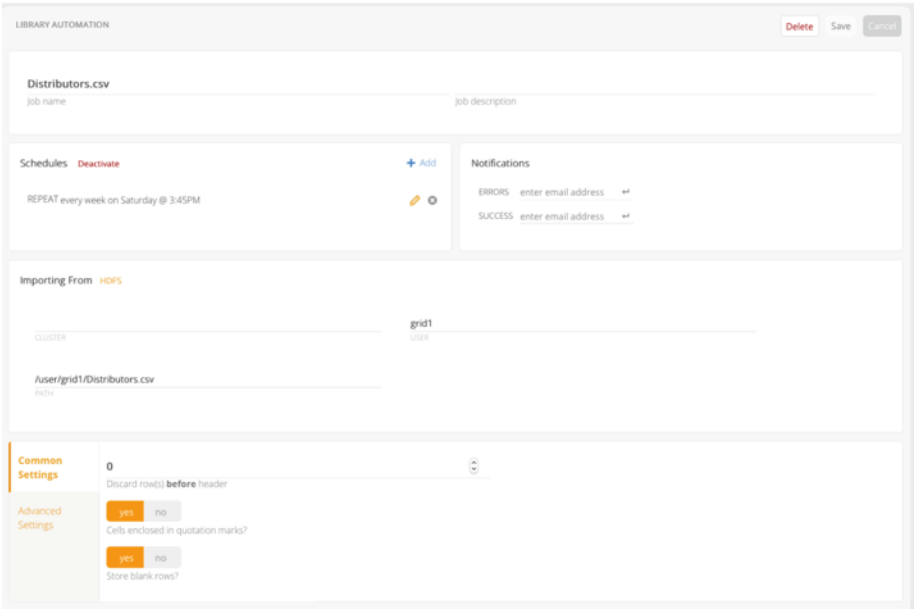
After you configure automation schedules for Data Library files and/or Projects, both are collectively referred to as automation "jobs".

Library Automation

To set up a Data Library file for automation:

1. Open the Data Library.
2. Locate the file you want to automate.
3. Hover over the row for the file and three icons appear. Click the timer icon:

4. The Automation page opens:



5. The application's in-line help provides the steps to walk you through setting up your Data Library file for automation.

Features and Considerations

- A Data Library file can be automated to run only once or on a recurring schedule that you specify.
    - The time you specify in the automation set up is when this file will be added to the queue for uploading, and not necessarily the start time for the automated import.
    - When configuring an automated job to run only once, be aware that your local computer's clock may not be precisely in synch with the web server that will process your job. For this reason, do not set a start time too near to the current time. If the web server's clock is ahead of your local computer's clock, the time you specify locally may have already passed for the web server and your job will not start.
- If you plan to automate a Project that will use this file for input: ensure a safe buffer of time for the file's update to finish before the Project's automated run begins.
- An automated job remains tied to the user account that originally scheduled it, and is the account under which the job will always run.
- Email notifications can be sent to notify users of either a successful upload into the Data Library or errors that have occurred. An Error email provides a link to the file's log file where you can determine the cause of the error(s). Note that recipients of the notification emails must have the required system permissions to view the automation results.

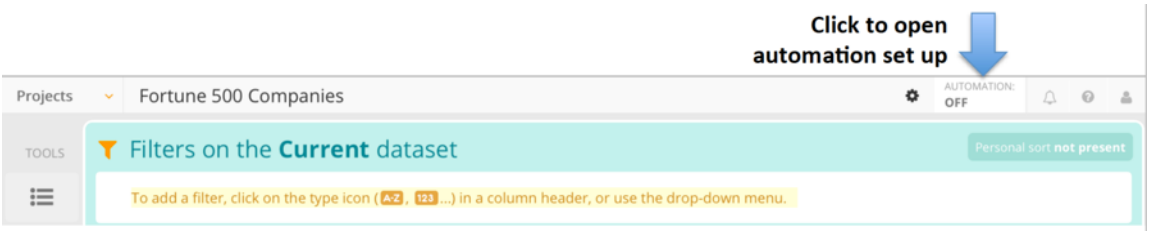
Project Automation

When you schedule a Project for automation, you set it up to automatically publish an AnswerSet in the Data Library based on the schedule and parameters you define. The AnswerSet can also be exported to external HDFS or exposed for query by JDBC (Impala or Hive) depending on your system's configuration.

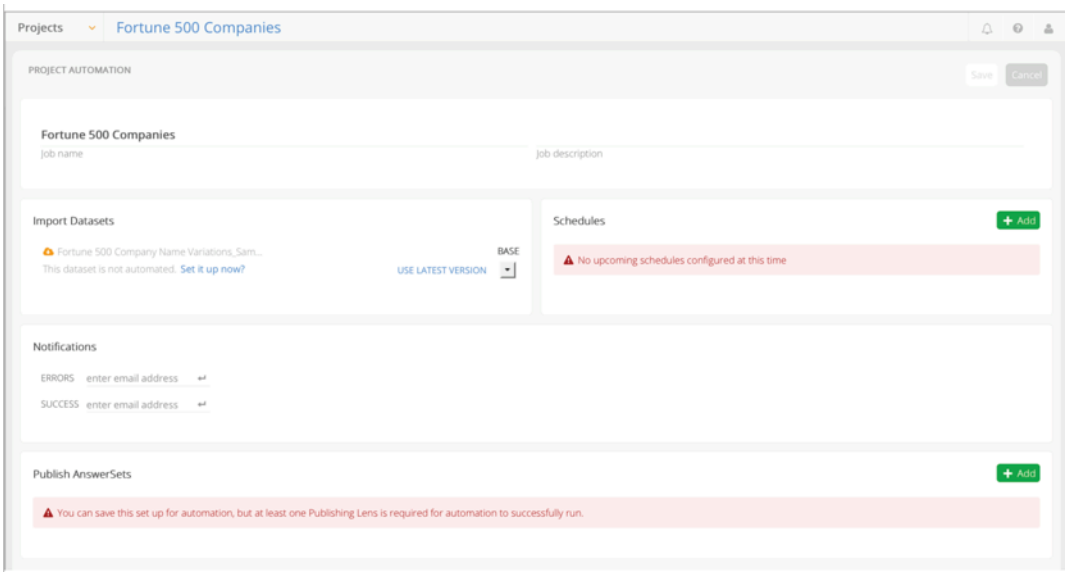
Note that Project Lenses are essential for Project Automation because they define the publishing points to use for your automated jobs. In order to automate a Project, you must have a Lens defined for each point in the Project where you want to publish data. You must have at least one Lens defined in your Project, otherwise no data can be published.

To set up a Project for automation:

1. Open the Project and click the automation status button:



2. The Automation page opens:



3. The application's online help provides the steps to walk you through setting up your Project for automation.

Features and Considerations

- A Project can be automated to run only once or on a recurring schedule that you specify.
    - The time you specify in the automation set up is when this Project will be added to the queue for publishing an AnswerSet, and not necessarily the publishing start time.
    - When configuring an automated job to run only once, be aware that your local computer's clock may not be precisely in synch with the web server that will process your job. For this reason, do not set a start time too near to the current time. If the web server's clock is ahead of your local computer's clock, the time you specify locally may have already passed for the web server and your job will not start.
- If a Project's automation depends on input from an automated Data Library file or an AnswerSet published from another automated Project: ensure a safe buffer of time for all input updates to finish before the automated run of the Project begins.
- An automated job remains tied to the user account that originally scheduled it, and is the account under which the job will always run.
- Email notifications can be sent to notify users when automated Projects finish updating or have errors. An Error email provides a link to the Project's log file where you can determine the cause of the error(s). Note that recipients of the notification emails must have the required system permissions to view the automation results.

Automation Dashboard

The Automation Dashboard provides history and details for all Data Library files and Projects that are set up to be automated. This is where you view the schedules for automated jobs and review their current status details. The dashboard is organized by:

The Automation Dashboard provides details and history for all Data Library files and Projects that are set up to be automated. This is where you:

- view your automation usage details.
- view the schedules for automated jobs.
- deactivate or delete job schedules.
- review the current status details for all jobs.

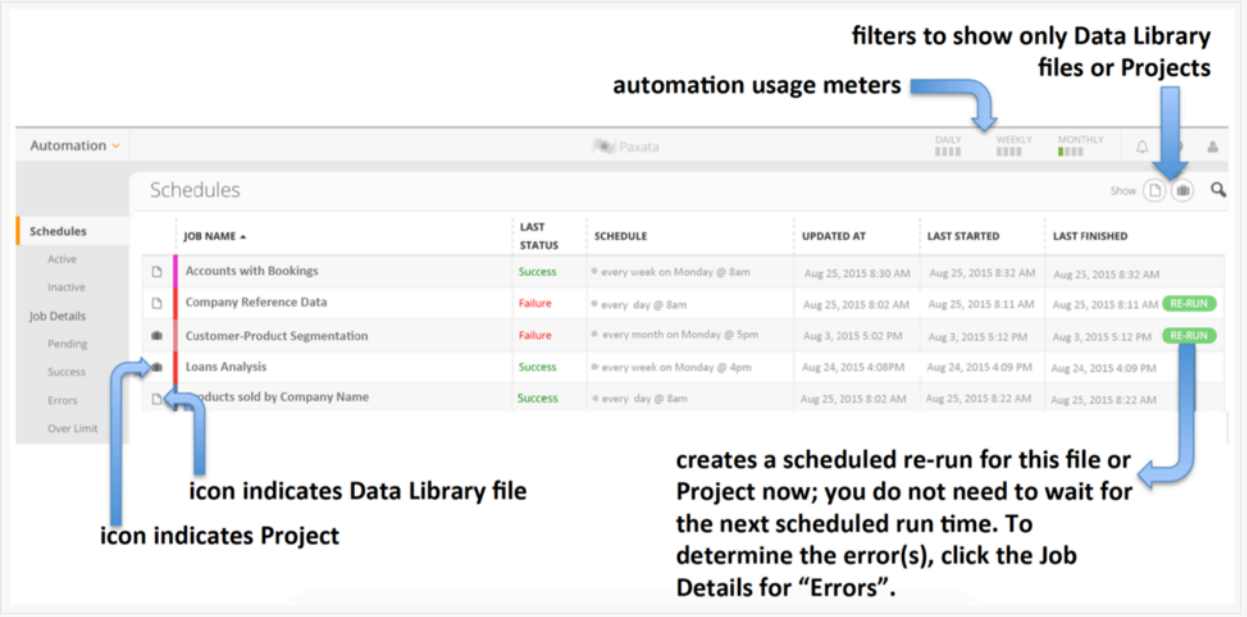
The dashboard is organized by **Schedules** and **Job Details**.

Schedules page

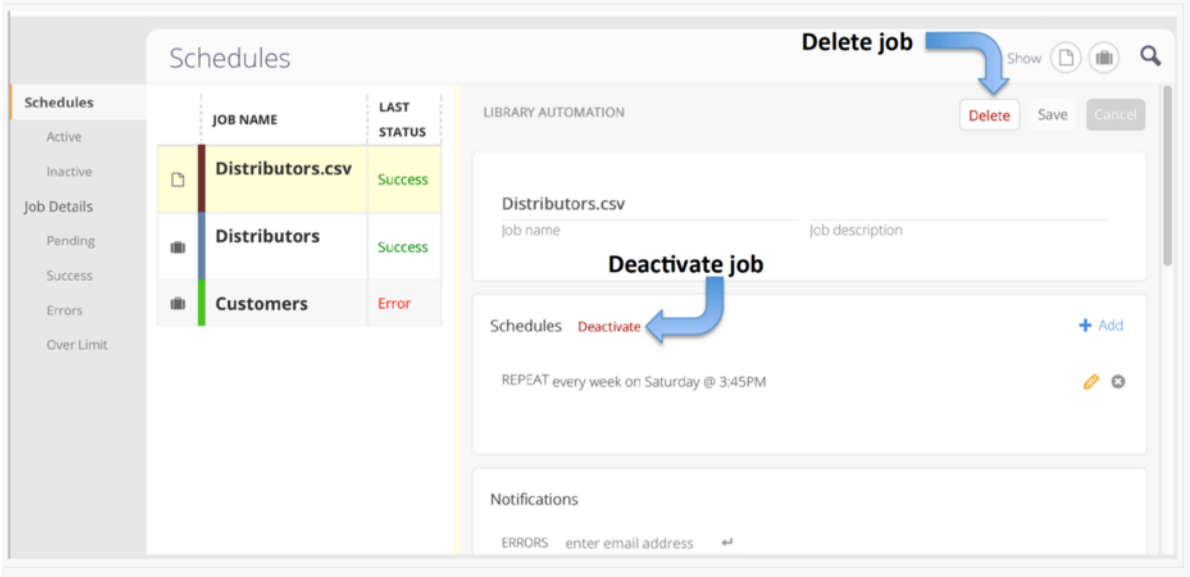
The Schedules page displays a list of all Data Library files and Projects that are currently configured for automation.

To view your automation usage details, mouse over the meters for additional information regarding the number of automated jobs you've already completed and the maximum number you can run for the day, week or month.

The Schedules page can also be filtered to display "Active" jobs or "Inactive" jobs that have had their automation schedules deactivated.



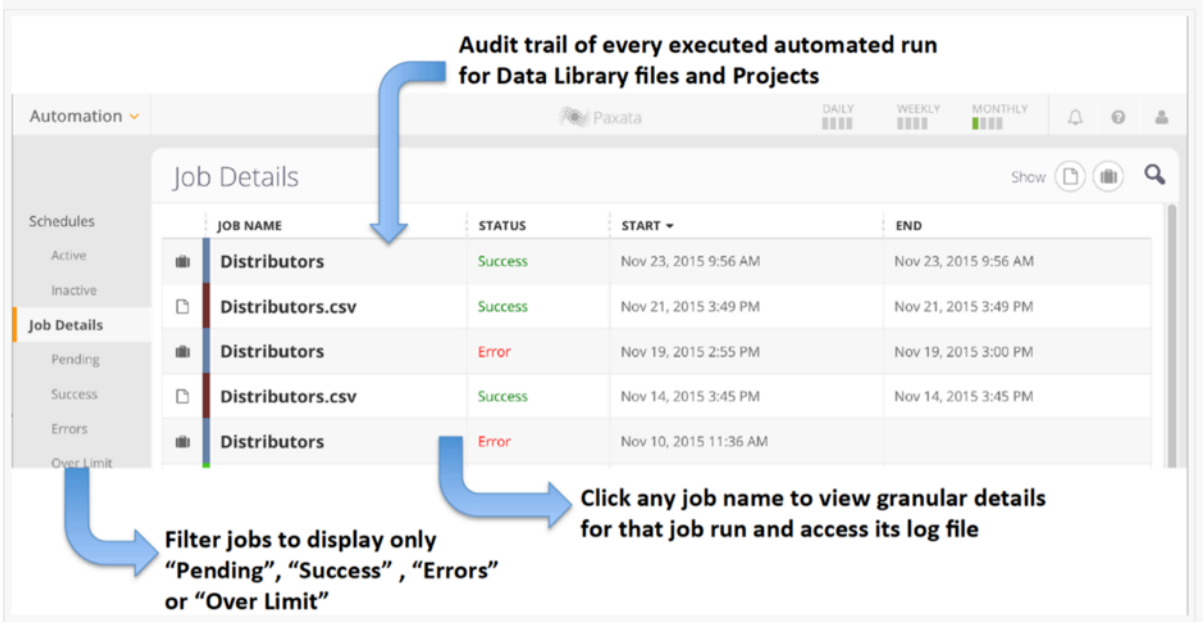
To deactivate, reactivate or delete a scheduled job: click the job's name to open its configuration page. Make and save the configuration change. Note that deleted jobs are not removed from the Job Details history.



Job Details

The Jobs Details page provides an audit trail for every executed automated run. From here you can also filter to view only the "Pending" jobs, jobs that have completed with "Success", jobs that have completed with "Errors", or "Over Limit" jobs that failed to run because automation limits were exceeded.

Note: only the Data Library files and Projects that you have permission to view are listed in the Dashboard.



The application's online help provides details of the actions you can perform in each area of the dashboard.

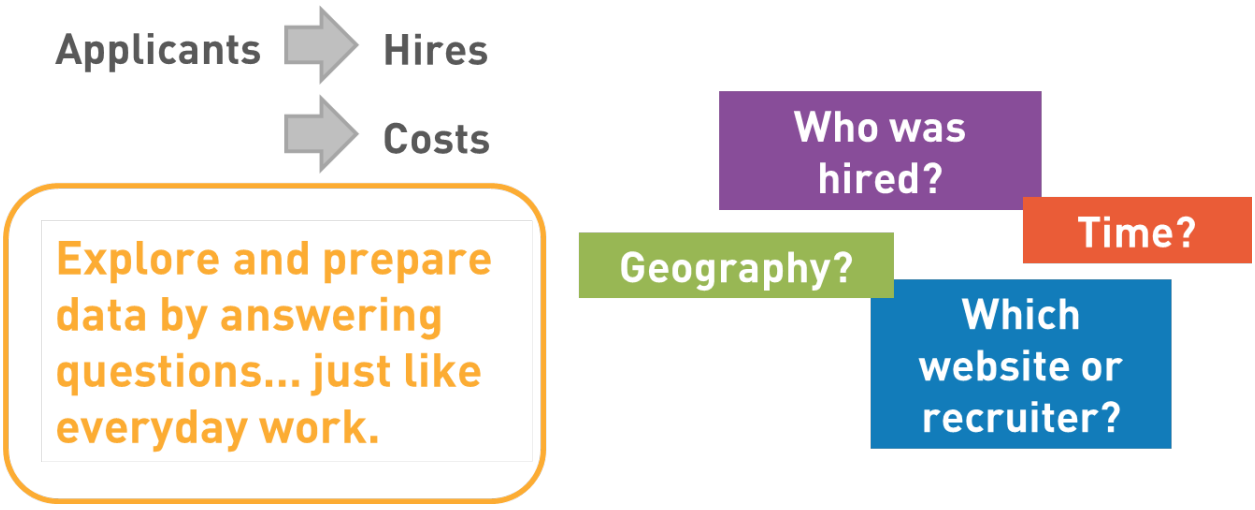
Putting It All Together

The steps shown below are designed to walk a beginning user through a simple use case with Paxata-supplied sample data. They show many of the most commonly used features and serve as an example of Adaptive Data Preparation™.

This example uses fictional human resources data related to hiring for the Acme Corp. There are three datasets involved:

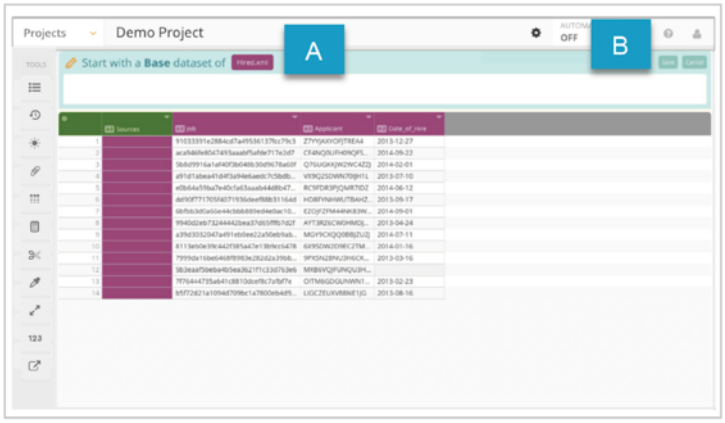
- Applicants.xlsx – This data is from the recruiting system. It contains applicant information, including jobs to which each applicant applied, and their unique identifiers. It came into Paxata as a Microsoft Excel® spreadsheet.
- Hired.xml – This file has been exported from the company's main human resources system, which is a separate system from the one used for recruiting. It maps applicants to the job for which they were hired and includes a hire date. The human resources system produced this file in an XML format.
- Job Postings.csv – Information from a third system shows where each job was posted, for which geographic region, and the time/costs involved. This file was uploaded as a comma-separated value (.csv) text file.

The goal of the use case is to produce an output file that shows a list of: all applicants who were hired, including their first and last names, source of each job posting, and the job titles for the accepted positions.



1. Select the *Hired.xml* [A] dataset from the Data Library. It contains the list of hires for the period in question.

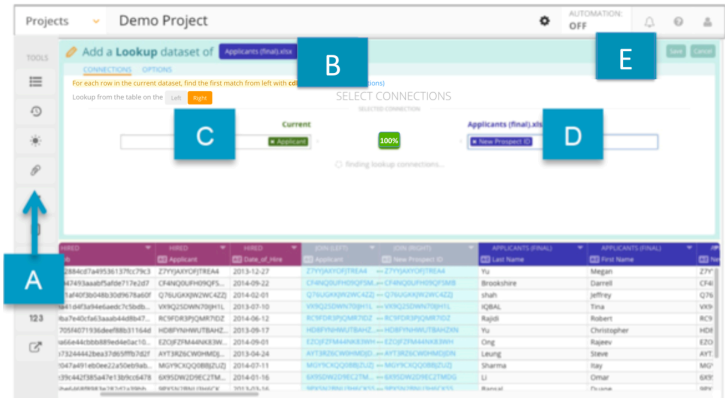
   Once the information is previewed in the Project, [B] click the “Save” button to commit it to the Project.



2. Because the *Hired.xml* dataset doesn’t contain any names, the first step is to bring in that information from the appropriate source.

   Click [A] the Lookup button and [B] select the *Applicants.xlsx* dataset from the Data Library. The [C] “Applicant” column from *Hired.xml* should be matched to the [D] “New Prospect ID” from the *Applicants.xlsx*.

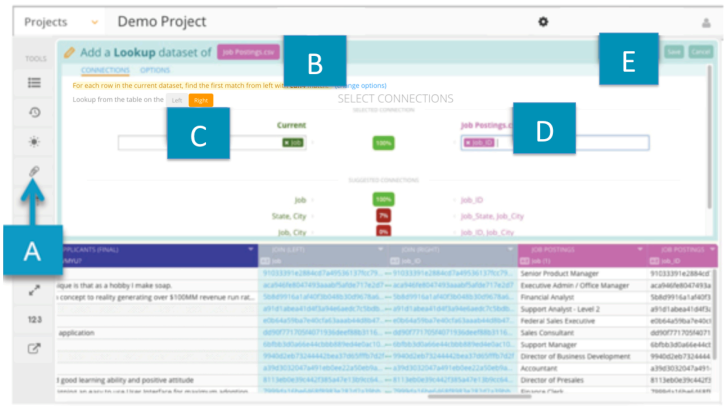
   After previewing the join, [E] click the “Save” button in the upper right-hand corner of the screen.



3. Immediately perform a second [A] Lookup operation. This one provides the job title from [B] the *Job Postings.csv* dataset.

   Once it has been selected, [C] join the “Job” column originally from *Hired.xml* to [D] the “Job_ID” column in the *Job Postings.csv* dataset.

   Once the Lookup has been previewed, [E] click “Save” to complete the operation.



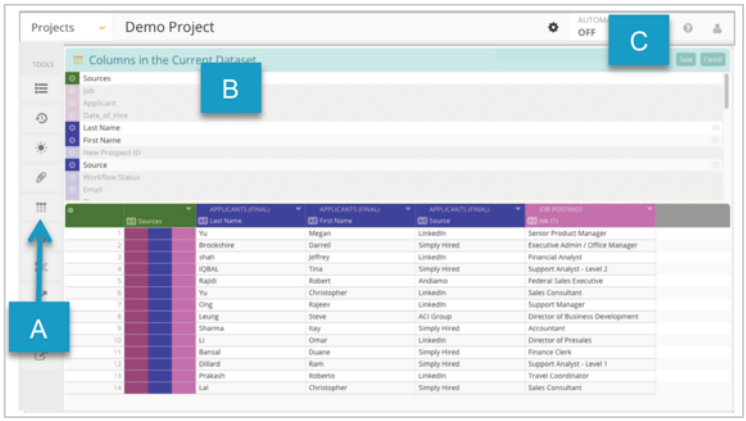
4.

Reduce the columns in the Project to just those you need for the answer set by [A] using the Manage Columns operation.

Scroll through the list of columns and [B] click each column to hide, leaving only the following columns visible:

- Last Name
- First Name
- Source
- Job (1)

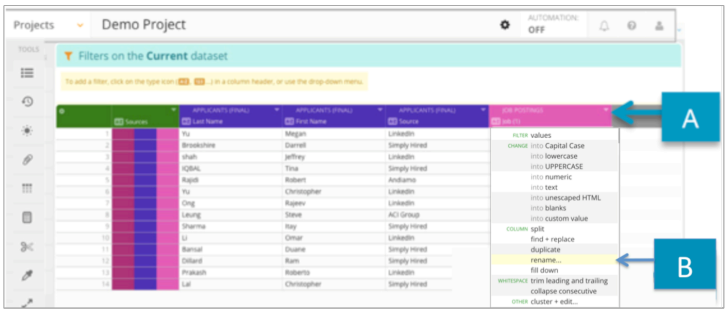
Note that you can use <Shift> and mouse clicks to select multiple columns at a time. Then [C] click "Save" to complete the operation.

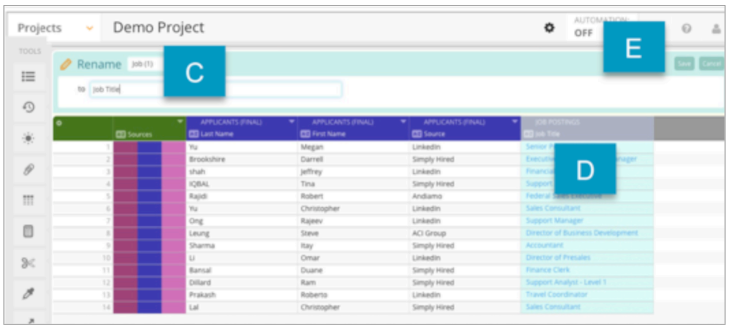


5.

Change the name of the "Job (1)" column to make it more descriptive. In [A] the column operations menu, [B] select **COLUMN rename**.

Enter [C] "Job Title" for the column header value and [D] preview the change below. Once complete, [E] click the "Save" button.



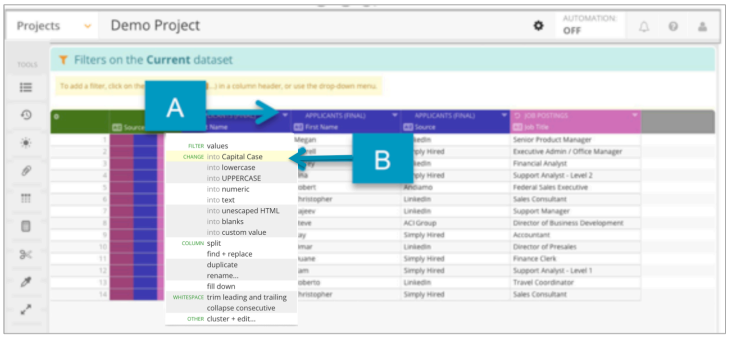


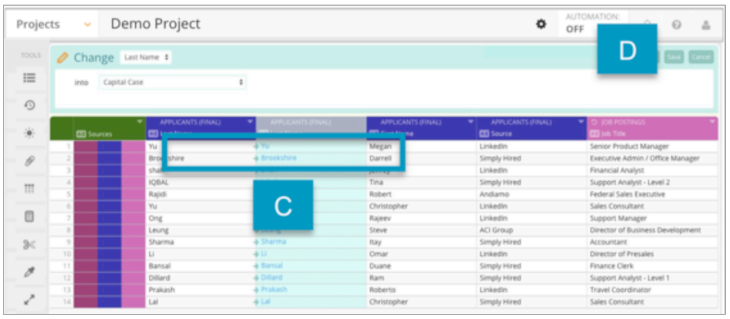
6.

If you examine the “Last Name” column, you’ll see that one name is lowercase (“shah”) and one is uppercase (“IQBAL”).

From [A] the column operations menu, [B] select transform into Capital Case.

Preview [C] the changes and then [D] click “Save” to save them to the Project.

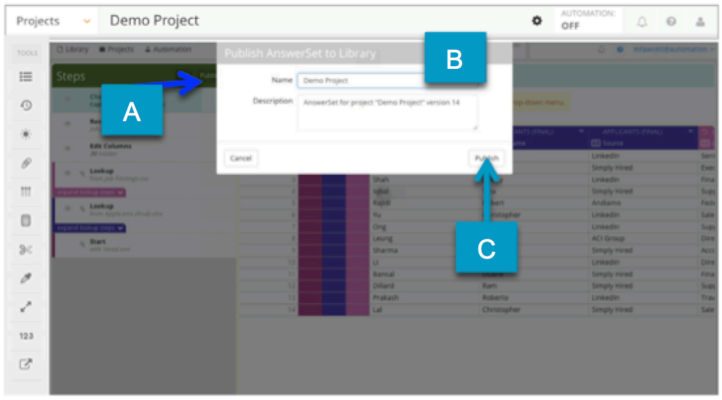




7.

Make the results of your work available to others [A] by clicking the Publish button.

Give [B] the answer set a helpful name and description. Click [C] “Publish” to make the information available in the Data Library.



Summary

Paxata empowers business teams to get from *raw* data to the *right* data… fast! It gives analysts an intuitive toolset to combine, clean and shape corporate and third-party information sources. The result: the right datasets that increase the accuracy and speed of decision-making.