# Prepare organizational data

This describes the value of organizational data for analysts and the steps of identifying, gathering, and structuring the data before you upload it.

To learn more about the nature and use of organizational data, see [Use organizational data for more effective analysis](#Xb36d027089c6501f1f9ac5469e7371f80784686). When you’re ready to start working with organizational data, the following sections describe how:

* [Identify trends that you want to analyze](#identify-trends-that-you-want-to-analyze) - Decide what trends you need to learn about to improve efficiency at work. From this, you can better choose what organizational data to use.
* [Know what data to include](#know-what-data-to-include) - A few data attributes are required, and many are optional. Among the optional ones, choose those that best serve your analytical purposes.
* [Get an export of organizational data](#get-an-export-of-organizational-data) - Have an admin export the HR data from your organization’s HR system. Optionally, include line-of-business data, if your analysis requires it.
* [Structure the organizational data](#structure-the-organizational-data) - For your data to validate successfully, you must first structure it correctly in the.csv file that you upload.
* [Upload the data to Workplace Analytics](#upload-the-data-to-workplace-analytics) - After your .csv file is ready, you upload it to Workplace Analytics where, after validation and processing, it becomes available for analysis.

## Use organizational data for more effective analysis

Organizational data is descriptive information about employees. After an admin uploads organizational data, Workplace Analytics combines it with Microsoft 365 data to provide detailed, actionable insights into the company’s communication and collaboration trends. An analyst can uncover these trends and use them to make more effective business decisions.

Here are examples of what analysts can do in Workplace Analytics after organizational data has been uploaded:

* Know how people communicate across job functions, department groups, and management hierarchies by enabling the grouping and filtering of descriptive attributes.
* Customize metrics to quantify group relationships, such as collaboration time between the marketing and sales groups.
* Make metrics calculations, such as insularity and redundancy.

Workplace Analytics automatically collects collaboration data from Microsoft 365. Analyzing just this data would create an incomplete picture; It’s the organizational data that you upload that provides analysis context. The following video illustrates these concepts:

### Video: Organizational data provides context

## Identify trends that you want to analyze

To know what organizational data to extract, you first need to decide what workplace trends you want to learn about. For example, in an upcoming analysis, you may want to examine collaboration across different employee segments, or groups. You must first define these groups, which you can do in various ways:

* Groups as defined by organizational data
* Groups made up of organizational hierarchy levels
* Groups made up of performance, engagement, or other line-of-business data

Defined groups can be used in the following examples of analyses:

### Collaboration between groups

A common analysis scenario is to find patterns of collaboration between different groups of employees. For example, you might want to know how much your product marketing team is talking to your sales team.

Attributes for segmenting populations can be helpful to consider in defining patterns of collaboration:

* Job family or role attributes, such as profession, function, discipline, and job code
* Organization, line of business, or cost center, such as HR, Finance, Sales, and Marketing
* Location attributes, such as city, state, country, and regions, as defined by your organization
* Attributes that describe the type of worker, such as remote, exempt/non-exempt, FTE/vendor, part time/full time, tenure in organization, and tenure in current role.

Most of these attributes are available within HR information systems.

### Hierarchical collaboration

It’s also common to seek patterns of collaboration behavior in reference to the hierarchy of your organization, as well as to quantify collaboration between managers and individual contributors, and between higher and lower levels and layers in the organization.

The following concepts are helpful in this analysis:

* **IC or manager** - Whether an employee is an individual contributor or a manager.
* **Organizational hierarchy** - For example, the names of all managers above the employee in that employee’s reporting structure; each manager can be stored as a separate attribute.
* **Layer** - For example, the position of the employee in the organizational hierarchy where layer 0 = the top leader in the company.
* **Span** - For example, the number of direct reports assigned to an employee.
* **Level** - For example, senior manager, VP, director, CVP, and so on.

Most of these attributes are also found in HR information systems.

### Collaboration, engagement, and outcome data

Finally, you might want to consider tying collaboration behavior patterns to employee engagement scores or other performance outcome data, such as sales-quota attainment or high/low performance ratings. This data is often found outside of traditional HR information systems, either in separate HR data repositories or in line-of-business systems.

## Know what data to include

To get full functionality from Workplace Analytics, you must supply several required attributes, as described in [Attribute reference](#attribute-reference). Additionally, you can supply up to 100 optional attributes to group and filter data in interesting and custom ways.

Examples of organizational data include: job family, job role, organization, line of business, cost center, location, region, layer, level, number of direct reports, and manager. This data is supplied to Workplace Analytics at the individual level, which means that these attributes provide context to each person in the dataset.

The following video describes which attributes are required and optional in your data upload:

### Video: What to include in the upload

### Which employees to include

At a minimum, include the organizational data for all employees who have Workplace Analytics licenses. It’s even better to include every person in your company as part of your data upload, even if you plan to gather collaboration data for only a subgroup, a specific target population within the company.

[!Important] If you upload data for everyone, you can analyze who everyone is collaborating with, even if they are outside your target population.

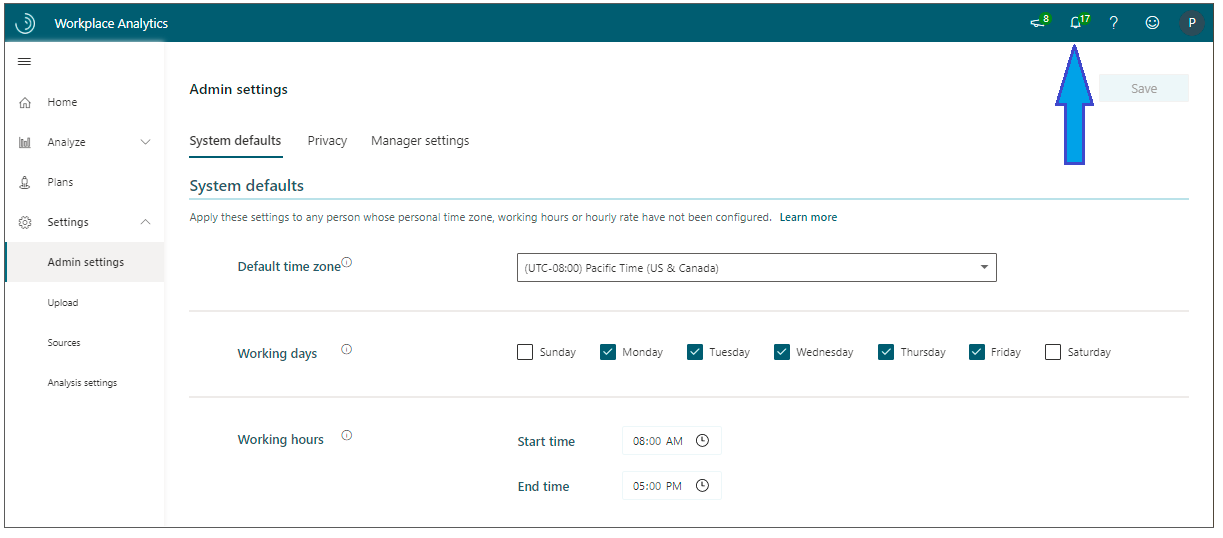
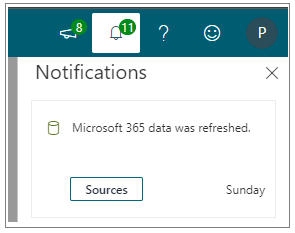
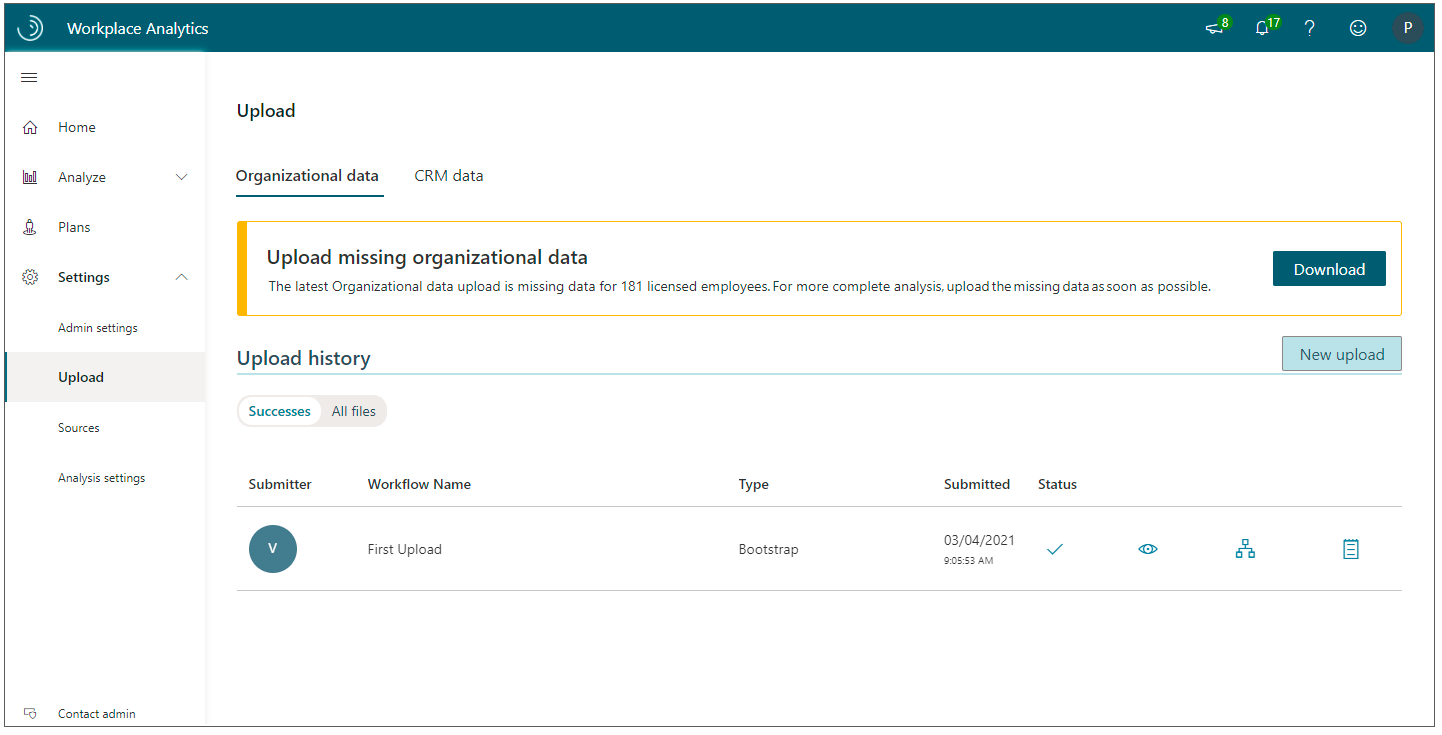
For example, if the people in Marketing communicate frequently with the people in Product Development, but Workplace Analytics has HR data only about the Marketing organization, you won’t be able to create reports to show how much time Marketing is spending with Product Development.

If you can’t include every person in your organization, the minimum to include is all people for whom collaboration data is being gathered. This minimum enables you to analyze collaboration patterns between groups within this population, but not between groups outside this population.

## Include all licensed employees

It’s the admin’s responsibility to maintain up-to-date and complete organizational data. In this task, “complete” means two things: include the right people and include the right attributes for those people.

Including the right people means, at a minimum, uploading organizational data for all licensed employees. If Workplace Analytics detects that data is missing for one or more licensed employees, it alerts admins in two ways:

* It sends an in-product notification to the admin. Alerts for these notifications appear on the bell icon:
* 
* Admin notification
* To see the notifications, select the bell icon. This opens the **Notifications** pane on the right side of the page:
* 
* Admin notification pane
* It shows a banner on the **Upload** > **Organizational data** page that reads **Upload missing organizational data**.
* 
* Banner for missing org data
* To respond to this warning, follow the steps in [Upload missing organizational data](#upload-missing-organizational-data).

### Upload missing organizational data

1. Select **Download**. This downloads a .csv file that contains the names of licensed employees whose organizational data is missing. For example:

* 
* Missing org data names

1. Open the .csv file.
2. Add the missing data for these employees. This means adding attributes (columns) that describe the employees in a way consistent with previous uploads. (See [Know what data to include](#know-what-data-to-include).)
3. Upload the file. (See [Upload organizational data (subsequent uploads)](upload-organizational-data.md).)

## Get an export of organizational data

Before you format and upload organizational data, you must get it from one or more sources. Your primary source is the team that manages your organization’s human resources (HR) information systems. This team will need to provide you with a data export of HR attributes for individual employees.

In addition, your analysts might need data about business outcomes. If so, you’ll need to contact line-of-business (LOB) owners who have access to data stores that contain this information. For example:

* Performance-review data for specific work groups
* Employee engagement scores captured by HR outside of HR information systems
* Sales or other quota-attainment data that provide additional views into performance
* Employee survey data

After you get this data, you must structure it for successful processing after uploading it to Workplace Analytics.

## Structure the organizational data

After you’ve identified what data to provide, you need to export it into the correct format to upload to Workplace Analytics. To start with, the data must be in a UTF-8 encoded .csv file and contain at least the set of required attributes for the population. For more information about saving a file in UTF-8 format, see [Solution](../tutorials/download-utf8-query-report.md#solution).

The file name must contain only alphanumeric characters (letters and numbers), with no spaces or special characters. For example, *FileName2.csv*.

The following video describes how to structure your organizational data file, including how to format the file, how to use the EffectiveDate field to reflect historical changes in your organization, which employees to include, and how to structure data that you add or update in subsequent uploads:

### Video: How to structure the organizational data file

### Required attributes

The following must be supplied with the exact column headers in the .csv upload, of which PersonId and ManagerId are not case sensitive, but the other text attributes are.

* PersonId
* EffectiveDate
* LevelDesignation (case sensitive)
* ManagerId
* Organization (case sensitive)

[!Note] The names of these required attributes are reserved, which means that you cannot use them as the names of [custom attributes](#custom-attributes).

### Reserved optional attributes

These are reserved column headers for attributes that are currently used only to filter and group data. As indicated, FunctionType and SupervisorIndicator are case sensitive.

| Attribute |Case sensitive? | If used, does a validation

threshold apply? | Can be used as the name of a

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| [custom attribute](#custom-attributes)? |  |  |  |  |
| FunctionType | Yes | Yes | No |  |
| HireDate | No | Yes | No |  |
| HourlyRate | No | Yes | No |  |
| Layer | No | Yes | No |  |
| SupervisorIndicator | Yes | Yes | No |  |
| TimeZone | No | Yes | No |  |
| GroupId | No | No | No |  |

### Custom attributes

These are any additional attributes you want to define to use in filtering and grouping data.

[!Note]

* The maximum number of total attributes allowed in the system is 105, which includes the five required attributes.
* All dates should be in the MM/DD/YYYY format.
* All numerical fields (such as the required attribute “HourlyRate”) must be in the “number” format and cannot contain commas or a dollar sign.

For more information, see [Attribute descriptions and data-coverage requirements](#X91022301e2c07aeed2fddc931021dcfda79f1b4) and [Video: What to include in the upload](#video-what-to-include-in-the-upload).

#### Example .csv export file

Here’s an example snippet of a valid .csv export file:

PersonId,EffectiveDate,HireDate,ManagerId,TimeZone,LevelDesignation,Organization,Layer,Area Emp1@contoso.com,12/1/2020,1/3/2014,Mgr1@contoso.com,Pacific Standard Time,5,Sales,8,Southeast Emp2@contoso.com,11/1/2020,1/3/2014,Mgr1@contoso.com,Pacific Standard Time,5,Sales,8,Southeast Emp3@contoso.com,12/1/2020,1/3/2014,Mgr2@contoso.com,Pacific Standard Time,4,Sales,7,Northeast Emp4@contoso.com,10/1/2020,8/15/2015,Mgr3@contoso.com,Pacific Standard Time,6,Sales,9,Midwest Emp5@contoso.com,11/1/2020,8/15/2015,Mgr3@contoso.com,Pacific Standard Time,6,Sales,9,Midwest Emp6@contoso.com,12/1/2020,8/15/2015,Mgr3@contoso.com,Pacific Standard Time,6,Sales,9,Midwest

For more information about attributes, see the [Attribute reference](#attribute-reference) section.

## Upload the data to Workplace Analytics

After you create a source .csv file, you can upload it to the Workplace Analytics service. If this is the first time that you will upload organizational data, see [Upload organizational data (first upload)](upload-organizational-data-1st.md). If this is not the first time, see [Upload organizational data (subsequent uploads)](upload-organizational-data.md).

After your data has been successfully uploaded, Workplace Analytics performs additional validation and processing to complete provisioning. If any problems occur, the Workplace Analytics team will contact your Workplace Analytics administrator.

### How often to upload organizational data

It is recommended that you upload HR data at least once a month to keep data fresh and analysis relevant. Soon after an HR upload has succeeded, the updated data becomes available in [Explore the stats](../use/explore-intro.md) and in [queries](../tutorials/query-basics.md).

#### Supplying data over a time period

By default, Workplace Analytics includes meeting and email data for measured employees for one year. Organizational data is provided to Workplace Analytics with an effective date associated with each row in the upload file, as described in [Attribute descriptions and data-coverage requirements](#X91022301e2c07aeed2fddc931021dcfda79f1b4) and in [Video: How to structure the organizational data file](#X45097c97550dc34915d89853e9c5b00fc78dfe7).

If you do a point-in-time export of organizational data from your HR information system as of the current date, you will get a picture of your employee population for that single point in time. Therefore, for the greatest data fidelity during provisioning, you should provide organizational data exports for each of the last 13 months. This can be supplied in a single file or in a sequence of files.

This means that for each measured employee, you would have 13 separate rows; each row would contain an effective date for each month in which data was pulled. If this is not possible, then you can provide one single point in time. In this case, the effective date should be set to the first day of the current month, one year back. For example, if provisioning occurred in October 2020, the effective date for all rows should be set to 10/1/2019.

The employee’s collaboration activity will be mapped to the most recent organizational data snapshot (based on EffectiveDate) that precedes the date of the collaboration activity. For a detailed example, see the video [Video: How to structure the organizational data file](#X45097c97550dc34915d89853e9c5b00fc78dfe7), starting at about 1:25.

## Attribute reference

This section contains information about the attributes that you use in the organizational data files uploaded to Workplace Analytics.

[Attribute descriptions and data-coverage requirements](#X91022301e2c07aeed2fddc931021dcfda79f1b4)

[Attribute notes and recommendations](#attribute-notes-and-recommendations)

### Attribute descriptions and data-coverage requirements

|  |  |  |
| --- | --- | --- |
| Attribute (column header) | Description of data / data validity | Data coverage requirements |

| PersonId |Unique identifier for the employee record. This identifier can be the employee’s primary SMTP address or email alias. It must be in a simplified format that contains no spaces. For example:

Allowed: person.name@xyz.com

Not allowed: <Name, Person> (person.name@xyz.com)

| Each row must contain a valid PersonId. Each upload file can have only ONE record with the same PersonID / EffectiveDate pair. | EffectiveDate |Date for which the given attribute value applies for the employee. The attribute applies until another record for the same attribute with a different effective date is specified. | Each row must contain a valid EffectiveDate. Each upload file can have only one record with the same PersonID / EffectiveDate pair. | | LevelDesignation | The employee’s level, which is represented as a string. This level is specific to your organization and can represent an employee’s experience or management level, or seniority within the organization. This data is needed to correctly calculate metrics for redundancy and insularity. | Each row must contain a LevelDesignation value. | | ManagerId | Unique identifier for the employee’s manager, which is needed to correctly calculate metrics for time spent with managers and their direct reports.This identifier can be the manager’s primary SMTP address or email alias. It must be in a simplified format that contains no spaces. For example:

Allowed: person.name@xyz.com

Not allowed: <Name, Person> (person.name@xyz.com)

| Each row must contain a valid ManagerId. | | Organization| The internal organization that the employee belongs to. An employee’s organization will be specific to your individual needs and could be identified by the leader of the organization, or by another naming convention. This data is needed to correctly calculate metrics for redundancy and insularity. | Each row must contain an organization value. | | FunctionType | The job function that the employee performs. This is specific to your organization. This data is used to filter and group reports, and for grouping of data in Explore the stats. | This attribute column is not required. If it is included, then each row must contain a function value.| HireDate| The date the employee began employment. This date determines the beginning date for calculating metrics of a measured employee. If an employee has multiple hire dates (for example: first hire date, most recent hire date), it is best to use the most recent hire date. | Each row should ideally contain a valid HireDate. If not included, metrics will be calculated from the start date of the data collection period.| |HourlyRate | The employee’s salary represented as an hourly rate in US dollars. **Notes**:

If the HR data only provides annual salaries, you’ll need to divide the employees’ salaries by 2080 to calculate their hourly rates in the upload (.csv) file before uploading it into Workplace Analytics.

The value can be formatted as a whole number, or include two decimal places, and cannot include any special characters, such as a dollar sign.

The value can represent pay only, or include the full value of benefits, as long as that choice is consistently applied for all employees.

This rate is used in calculations and can be used to filter and group employees.

If the upload doesn’t include an hourly rate for an employee, Workplace Analytics uses a default HourlyRate of $75 for calculations and metrics.

You can change the default rate in [Admin settings](../use/admin-settings.md). If you change the default, this change applies retroactively to anyone without an effective hourly rate for the next scheduled refresh of your organizational (HR) or Microsoft 365 collaboration data. For more information, see [System defaults](../use/system-defaults.md).|This attribute column is not required. If it is included, then each row must contain a floating point or integer value with no special characters (such as a dollar sign).| |Layer | The place where the employee is within the organizational hierarchy. Layer is represented as an integer and expressed as the distance the employee is from the top leader of the organization. For example, the CEO, is at layer 0. This data is used to filter and group reports, and for grouping of data in [Explore the stats](../use/explore-intro.md) features. | This attribute column is not required. If it is included, then each row must contain an integer value.| |SupervisorIndicator | Use this attribute to view the habits of people managers or influencers in your organization in Power BI visualizations. It powers the Overview table, the Generated Workload charts that are generated when you use the Manager Impact [Power BI template](../tutorials/power-bi-templates.md), and the [Influence insights Power BI dashboard](../tutorials/pbi-influence-db.md).

This attribute indicates the manager status of each employee as IC (individual contributor), Mngr (manager), or Mngr+ (manager of managers); however, note that if different nomenclature is used in your file, you must update the Power BI chart filters accordingly. If you include SupervisorIndicator, you must also include the values **IC**, **Mngr**, or **Mngr+** in your organizational data. | This attribute is required if you want to use either of the following:

the Manager impact dashboard in Power BI

the template that you use to create the [Influence insights Power BI dashboard](../tutorials/pbi-influence-db.md)

| |TimeZone |Time zone in which the employee performs work. This must be one of the time zones in [Time zones for Workplace Analytics](../use/timezones-for-workplace-analytics.md). If you do not have a time zone available for each employee, the system will use the default, which is Pacific Standard Time. | This attribute column is not required. If it is not included, the default time zone will be used.| |Any user-defined columns | Additional columns can represent any data that you want to use in queries to group and filter employee records. | No coverage requirements. |

### Attribute notes and recommendations

#### Some attributes exist only for a subset of the population

When choosing attributes to include, some attribute values might be populated for one organization but not others. For example, if the upload includes sales quota-attainment data that only applies to your sales organization, you cannot use this data for filtering and grouping employees outside of sales.

#### Use only allowed time zones

The default time zone for Workplace Analytics is Pacific Standard Time (PST). See [Time zones for Workplace Analytics](../Use/Timezones-for-workplace-analytics.md) for a complete list of the times zones that you can use.

#### Too many unique values

Sometimes an attribute has too many unique values to use for grouping and filtering. For example, if a job function or code is too narrowly defined, it might not give you a useful view of the overall group. If an attribute has hundreds of unique values that result in a small population group per value, the attribute might not be useful.

#### Too few unique values

Conversely, sometimes an attribute is too broadly defined for useful filtering. For example, if your organization resides entirely in the United States and your HR records per employee contain a country code that always equals US, that attribute won’t be useful.

#### Redundant attributes

Some attributes might represent the same data and provide unnecessary redundant data for analysis. For example, HR data could contain both a cost center ID and a cost center name for an employee. Because both represent the same information in a slightly different format, you’ll want to include only the one with the more “user friendly” name.

#### Line-of-business data

Unlike HR data, for line-of-business data, you might not need to include every person in your company as part of your data upload. Knowing the scenarios you want to analyze will help you to decide.

For example, suppose you want to compare collaboration patterns between employees in the Sales organization who have high engagement as compared to those who have low engagement. Although you will want HR data for all employees so you can characterize broader collaboration patterns, you only need engagement score data for employees in the Sales organization, because you will use the score values to group and filter specific report outputs.

### Use only valid values and formats

When any data row or column has an invalid value for any attribute, the entire upload will fail until the source file is fixed (or the mapping changes the validation type of the attribute in a way that makes the value valid).

[!INCLUDE [Valid values and formats](../includes/org-data-upload-tips.md)]