

# TIBCO® Health Essentials Cloud

## Data Prep User's Guide

*Release 4.0*

*January 2022*

*DN3509485.0122*





# Contents

---

<b>1. Introducing TIBCO® Health Essentials Cloud .....</b>	<b>5</b>
Input Data Sets .....	5
Custom Relationship File.....	8
Data Prep Guidelines .....	8
Primary Identifier .....	9
Alternative Primary Identifier .....	11
Master Identifier .....	11
Alternate Master Identifier .....	12
Coded Data .....	12
Reference Between Data Sets .....	13
File Naming Convention .....	13
File Layout Specifications .....	14
Data Validation Guidelines .....	14
Data Processing Guidelines .....	17
Scheduling Analytics Jobs .....	17



## Introducing TIBCO® Health Essentials Cloud

---

Electronic health data may come from a wide variety of sources, including electronic health records (EHRs), administrative claims processing systems, diagnostic laboratory systems, imaging systems, medical devices, and consumer devices. Integrating data sources can take many forms depending on the type(s) of data and the analytics required to support the business goals of a healthcare organization. In some cases, data analysts will work directly with individual systems to integrate or link to data, while, in other cases, data analysts may work with sources in which the data is already aggregated and standardized, such as clinical data warehouses and health information exchanges. The purpose of this guide is to describe the technical approaches and key questions to consider that may be used to incorporate and prepare this data into a set of input data files supported by TIBCO® Health Essentials Cloud. It is intended for data analysts and data integrators.





### In this chapter:




- [Input Data Sets](#)
  - [Data Prep Guidelines](#)
  - [Primary Identifier](#)
  - [Alternative Primary Identifier](#)
  - [Master Identifier](#)
  - [Alternate Master Identifier](#)
  - [Coded Data](#)
  - [Reference Between Data Sets](#)
  - [File Naming Convention](#)
  - [File Layout Specifications](#)
  - [Data Validation Guidelines](#)
  - [Data Processing Guidelines](#)
  - [Scheduling Analytics Jobs](#)
- 


## Input Data Sets

TIBCO® Health Essentials Cloud (HEC) provides a set of data file specifications that allow data analysts to integrate data from different health systems into a set of consolidated data files that can be uploaded, validated, and processed to Analytics data marts, powering business applications that satisfy a healthcare entity's analytics and reporting needs. The following table describes the set of input data files supported by HEC.

For more information on these data sets, see the corresponding input file specifications that are attached to this PDF for your reference.

Data File	Description
<p>Patient</p> 	<p>A patient is an individual who is a recipient of healthcare services. This file specification is used for collecting patient information, for example, demographics, identifiers, addresses, and preferred providers.</p> <p><b>Note:</b> The data in this file may have references to records in the Provider file.</p>
<p>Provider</p> 	<p>A provider is an individual or an organization that participates in delivery of care services or products. This file is used for collecting care provider information, for example, demographics, identifiers, addresses, and provider specialties.</p>
<p>Facility</p> 	<p>A facility represents a physical plot of land that may be owned, bought, sold, or leased. This file is used to capture base level facility information such as physical location, owner, contact methods, and name.</p> <p>Some examples of facilities are hospital, parking garage, and administrative building.</p> <p><b>Note:</b> The data in this file has references to records in the Facility file.</p>
<p>Facility Location</p> 	<p>A facility location represents a physical place where healthcare services are provided. A Facility Location is always defined within the context of a physical Facility.</p> <p>Some examples of facility locations are floor, ward, room, and bed.</p> <p><b>Note:</b> The data in this file has references to records in the Facility file.</p>

Data File	Description
Encounter 	<p>An encounter is an interaction between a patient and healthcare provider or providers for the purpose of providing healthcare services or assessing the health status of a patient. A patient encounter is further characterized by the setting in which it takes place, such as ambulatory, emergency, home health, inpatient, and virtual encounters. During the encounter, the patient may move from practitioner to practitioner and location to location and each of these events are captured through individual clinical events.</p> <p><b>Note:</b> The data in this file may have references to records in the Patient and Facility files.</p>
Clinical Events 	<p>A clinical event represents clinical facts reported or recorded during a clinical encounter involving a patient and a healthcare provider.</p> <p>Supported types of clinical facts include Condition (Diagnosis), Procedure, Immunization, Medication, Laboratory Observation, and Vital Sign Observation.</p> <p><b>Note:</b> The data in this file may have references to records in the Encounter, Facility Location, and Provider files.</p>
Customer Code Sets 	<p>This file specification is used by a customer to provide codes that are either licensed by the customer, such as AMA CPT codes, or any other proprietary codes needed to process the data in Omni.</p>

Data File	Description
<p>Custom Relationships</p> 	<p>This file specification is used by a customer to define relationships, thresholds, measures, and so on, that can be used for reporting and analytics. The data in this file is directly loaded into a relationship table in the customer's data mart and does not go through Omni workorder processing.</p> <p>Example use cases are:</p> <ol style="list-style-type: none"> <li>1. Records in this file can be used to indicate a Provider to Patient relationship. A user (physician) can run a report of attributed patients and can only access the Patient Journey reports for those individual patients.</li> <li>2. Define operational goals, such as Outpatient performance targets, for example, target volume of well visits or sick visits.</li> </ol>

### Custom Relationship File

This is a special data set that allows customers to upload any custom relationships they need to access in the TIBCO WebFOCUS® analytics mart without having to process through Omni workflows.

Typical uses for this data set:

- Setup thresholds for target metrics that are used in various reports
- Setup provider-patient attribution list for setting up access to Patient Journey app

**Note:** Whenever a user uploads this data set for processing, it replaces the previously loaded data. Users must exercise caution and follow these best practices:

1. Archive a copy of every custom relation data set that is uploaded for processing.
2. Verify that the new custom data set retains previously loaded data that is not intended to be altered or deleted.

### Data Prep Guidelines

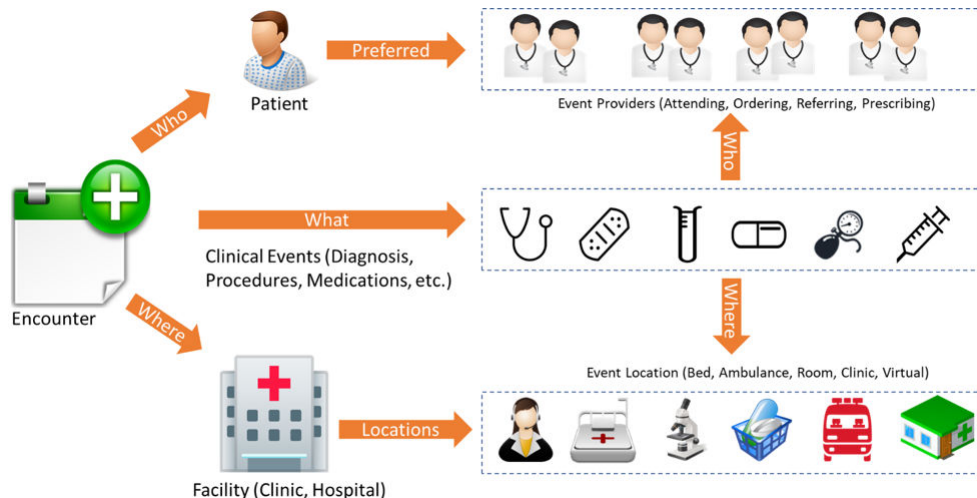
Analysts preparing data files for TIBCO® Health Essentials Cloud (HEC) should refer to the following relationship diagram to improve their understanding of the different data sets needed in HEC.

1. Identify the list of source systems that capture patient information and records of care delivered to those patients.



2. Create a list of Patients who have received care.
3. Create a list of Clinical Encounters with information such as who received care (Patient), where the encounter occurred (Facility), and the duration of that encounter.
4. Create a list of Clinical Events generated during each encounter.
5. Create a list of Providers who have participated in delivering care to the Patients in an encounter.
6. Create a list of Locations within a Facility where care was delivered during an encounter.

Data analysts need to identify and access the different data sources that capture different aspects of care provided to Patients. Each of the Patients in the Patient Registry would have participated in a number of Clinical Encounters. Each Encounter will be associated with a collection of Clinical Events. These Clinical Events will have a reference to participating providers and locations where the event occurred, as shown in the following image.



## Primary Identifier

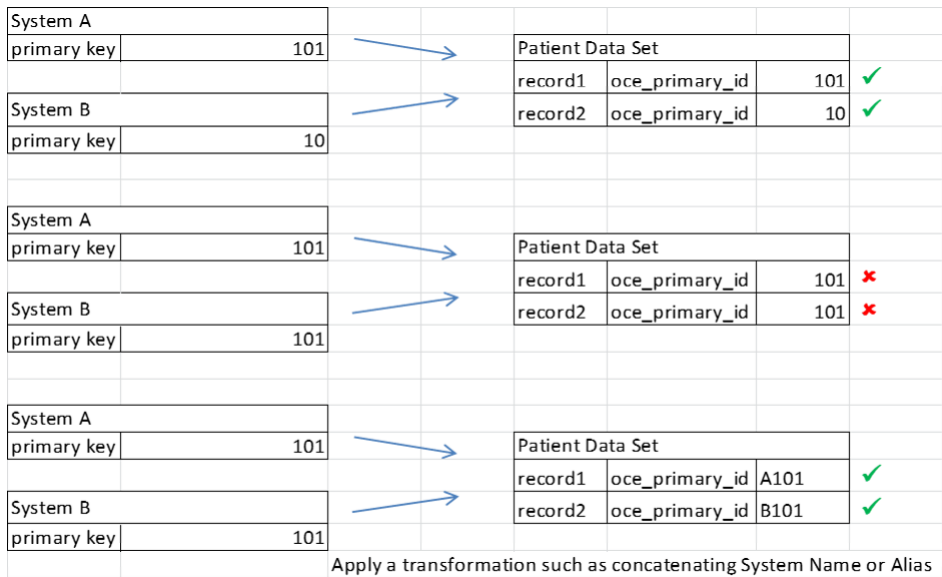
The Primary Identifier (`oce_primary_id`) field provides the unique identifier of a record in the data set.

## Constraints

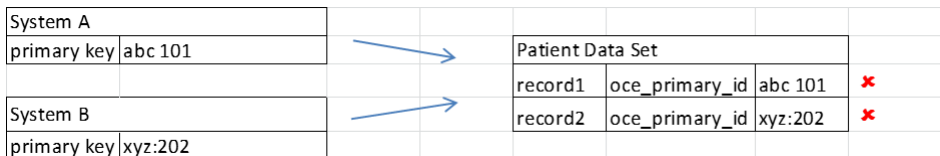
The following are constraints of the Primary Identifier (oce\_primary\_id) field.

1. **Cannot be duplicate.** If there are duplicate records in a data set identified by the same *oce\_primary\_id* value, all the duplicate records will be removed from the data set. The validation report generated for the data set will contain record numbers for the rows that were identified as duplicates. Duplicate records are removed from a data set when loading the data into OnRamp tables.

An alternative is to identify a different attribute for patient data in the source system that uniquely identifies the record across all systems or, apply a simple transformation to add the source system name or alias to uniquely identify a record from that system, as shown in the following image.



2. **Cannot have embedded spaces or colon (:).** The unique identifier of a record in a data set cannot have embedded spaces in the value or a colon (:) separated set of values. If the source system has identifiers that have embedded spaces or colons, then use the *oce\_primary\_id\_alt* field, as shown in the following image.



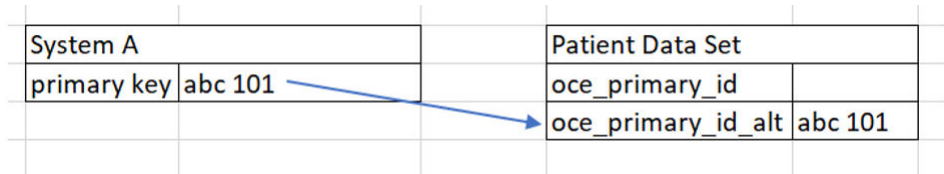
## Alternative Primary Identifier

Use the Alternative Primary Identifier (`oce_primary_id_alt`) field only when the primary record identifier in the source system has embedded spaces or colons (:). If the source system cannot generate a unique identifier value that satisfies the constraints defined for `oce_primary_id`, then use this field to provide the alternate that can be used by Omni to generate a unique identifier for the record in a data set.

### Constraints

The following are constraints of the Alternative Primary Identifier (`oce_primary_id_alt`) field.

1. **Cannot be duplicate.** If the `oce_primary_id` field is empty and there are duplicate records in a data set identified by the same `oce_primary_id_alt` value, all the duplicate records will be removed from the data set. The validation report generated for the data set will contain record numbers for the rows that were identified as duplicates. Duplicate records are removed from a data set when loading the data into OnRamp tables. Omni will autogenerate the `oce_primary_id` field using the value provided in the `oce_primary_id_alt` field, as shown in the following image.



## Master Identifier

Use the Master Identifier (`oce_master_id`) field to provide the master identifier (golden record identifier or unique identifier for an entity across different systems). This field exists only in the mastered data sets, for example, Patient, Provider, Facility, and Facility Location. Examples of values that can be used for the `oce_master_id` field are shown in the following image.

data set	oce_master_id
Patient	ssn or empi
Provider	npi or empi
Facility	federal tax id

### Constraints

The following are the constraints of the Master Identifier (*oce\_master\_id*) field.

1. **Cannot have embedded spaces or colon (:).** The master record identifier cannot have embedded spaces in the value or a colon (:) separated set of values. If the source system has master record identifiers that have embedded spaces or colons then use the *oce\_master\_id\_alt* field.

## Alternate Master Identifier

Use the Alternate Master Identifier (*oce\_master\_id*) field to provide alternative values for master identifier (golden record identifier or unique identifier for an entity across different systems) if the master identifier contains embedded spaces or colons (:). Omni will use the values in this field to autogenerate the value for the *oce\_master\_id* field.

## Coded Data

There are several fields in different data sets that are defined as coded elements. Codes are used for encoding data, such as tables of terms, medical concepts, medical diagnoses, or medical procedures. Coding prevents inaccuracies in descriptions of diseases, injuries, healthcare procedures, and so on, from impacting the outcome of analysis of those facts. There are two kinds of coded fields defined in the data specifications.

1. **Coded fields with the code list available in the specifications.** These are coded fields where a list of allowed codes is provided in the specifications. Any record with a coded field value outside of the allowed list of values will generate a validation error, and depending on the severity of that error, the record may get accepted or rejected as a fatal error.

For example, the Gender code in the Patient data set is a required field, as shown in the following image. If a record does not include one of the allowed values from code set 0001, that record will generate a fatal error during data validation and data processing steps.

<i>oce_gender_code</i>	coded	1	Required	This is a code classifying the gender that the patient is considered to have for administration and record keeping purposes.	0001
------------------------	-------	---	----------	--	------

The following table shows the allowed values for Gender code.

0001	A	Ambiguous
0001	F	Female
0001	M	Male
0001	N	Not applicable
0001	O	Other
0001	U	Unknown

2. **Coded fields with a standard industry code set binding.** These are coded fields where the list of allowed codes is omitted from the specifications because they are standard industry terminologies, such as ICD10, HCPCS, and NDC.

For example, fields for diagnosis codes in the Clinical Events data set reference the standard ICD10 and SNOMED codes, as shown in the following image. There is no link to the code set in the data specification because it is assumed that these standard codes are available in the source system generating the data for the clinical events.

oce_diag_icd10_code	coded	8 xxx.xxxx	Required	The International Classification of Diseases (Tenth Revision) code representing the clinical condition. This code should have a decimal point if it is more than 3 characters long (category.sub-category).	ICD10-CM
oce_diag_snomed_code	coded	18	Preferred	The Systematized Nomenclature of Medicine - Clinical Terms code representing the clinical condition.	SNOMED

**Important:** The Current Procedural Terminology (CPT) and HIPAA Provider Taxonomy standard code sets are extensively used in TIBCO® Health Essentials Cloud (HEC) for analytics, but are not provisioned in the HEC environment because of copyright and distribution restrictions by AMA and WPC. Customers need to upload these two code sets using the Customer Code Set file specifications.

## Reference Between Data Sets

Records in a data set can have references to records in other data sets.

For example:

- Reference to a preferred provider record identifier in the Patient data set.
- Reference to a patient record identifier in the Encounter data set.

These references are provided using either the `_id` or `_id_alt` fields and must match the `oce_primary_id` or `oce_primary_id_alt` values for the corresponding records in that data set, as shown in the following image.

System A (Patient Data)			Patient Data Set	
primary key	pat101	→	oce_primary_id	pat101
System A (Encounter Data)			Encounter Data Set	
primary key	enc22	→	oce_primary_id	enc22
patient foreign key	pat101	→	.....	
		→	oce_patient_id	pat101

## File Naming Convention

Users cannot upload a file with the same name within a duration of 10 minutes.

Accepted file name extensions are:

- .gz for compressed files.
- A file with any other extension is treated as a text file.

**Note:** It is recommended to add a date time stamp to the file name to make it unique.

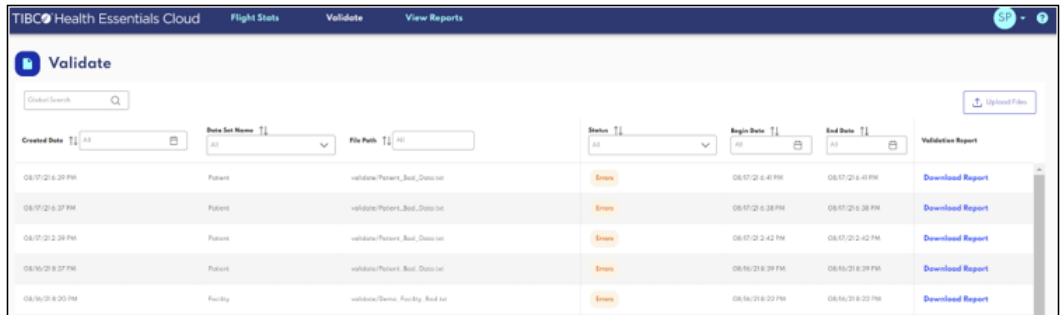
## File Layout Specifications

Records in a TIBCO® Health Essentials file must comply with the following specifications:

1. Must be pipe delimited.
2. Column headers are required.
3. All columns must be present and column data should match the column header. The order of columns is not important.

## Data Validation Guidelines

A user with Data Analyst privileges can log in to the TIBCO® Health Essentials Cloud Console to access the Validate tab and upload files for analyzing the data against validations rules. The user can then access and download the report generated from the validation process and take necessary action to fix errors reported in the validation output report. The Validate tab is shown in the following image.



The following table lists and describes the validation error codes.

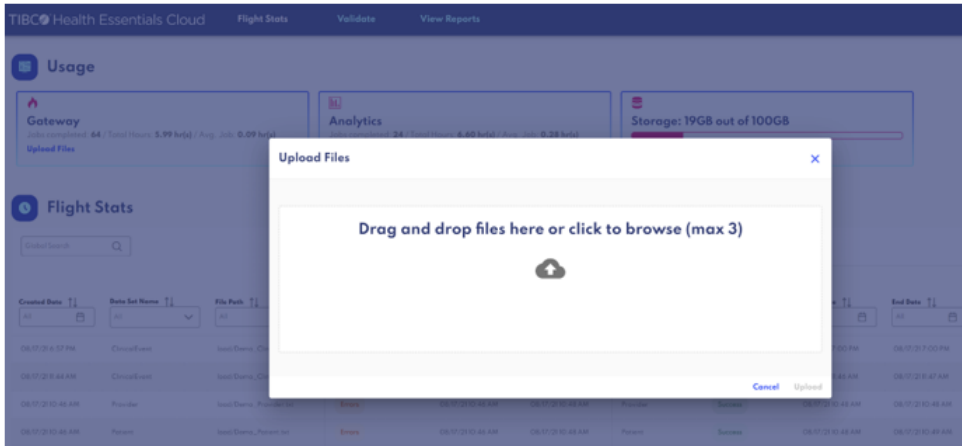
Reason Code	Description
TRIMMED	<p>Record has a field value with leading or trailing spaces.</p> <p>Field value is trimmed to remove leading and trailing spaces.</p>
INVALID_DUPLICATE	<p>Record has a duplicate primary identifier.</p> <p>Record is rejected.</p>
INVALID_FATAL	<p>Record has invalid values for a critical field.</p> <p>Record is rejected.</p>
TRUNCATED	<p>Record has a field value that exceeds the field length.</p> <p>Field value is truncated.</p>
TRUNCATED,INVALID_FATAL	<p>Record has a field value that exceeds the field length.</p> <p>Field value is truncated.</p> <p>Record has a truncated field value that is invalid.</p> <p>Record is rejected.</p>
INVALID_REPLACED	<p>Record has a field value that is invalid.</p> <p>Field value is replaced with a default value.</p>
TRUNCATED,INVALID_REPLACED	<p>Record has a field value that exceeds the field length.</p> <p>Field value is truncated.</p> <p>Record has a truncated field value that is invalid.</p> <p>Truncated field value is replaced with a default value.</p>
INVALID_ACCEPTED	<p>Record has a field value that is invalid.</p> <p>Invalid field value is accepted, no further action is taken.</p>

Reason Code	Description
TRUNCATED,INVALID_ACCEPTED	<p>Record has a field value that exceeds the field length.</p> <p>Field value is truncated.</p> <p>Record has a truncated field value that is invalid.</p> <p>Invalid truncated field value is accepted, no further action is taken.</p>
INVALID_NULLIFIED	<p>Record has a field value that is invalid</p> <p>Field value is replaced with Null.</p>
TRUNCATED,INVALID_NULLIFIED	<p>Record has a field value that exceeds the field length.</p> <p>Field value is truncated.</p> <p>Record has a truncated field value that is invalid.</p> <p>Invalid truncated field value is replaced with Null.</p>



## Data Processing Guidelines

A user with Operations Analyst privileges can upload data through the Gateway Console up to three files at a time. Once uploaded, a new entry for each file appears on the Flight Stats screen that shows the processing status for that data set.



## Scheduling Analytics Jobs

Analytics jobs can be run on-demand or on a predefined schedule.

The default schedule that is setup for the analytics job is:

- Frequency (Once a week)
- Day of week (Friday)
- Time of day (11pm US Eastern Standard Time)

If your business requires a different schedule, create a ticket on [TIBCO® Support](#) and the TIBCO® Health Essentials Cloud Support team will implement the change within two business days.

