

# Peste Bubónica en Iquique, 1903: Transcripción de datos no tabulares usando DataScribe

por Hernán Adasme

<b>Introducción</b>	<b>2</b>
<b>Contexto histórico</b>	<b>2</b>
<b>Acerca de los datos</b>	<b>4</b>
<b>Configuración del proyecto DataScribe parte 1: ítems en la instalación de Omeka S</b>	<b>5</b>
Aregar los ítems en Omeka S	5
Creación de conjuntos de ítems	6
<b>Configuración del proyecto DataScribe, parte 2: creación de un proyecto DataScribe</b>	<b>7</b>
Creando una Peste en Iquique 1903 Proyecto DataScribe	7
Construyendo el Formulario de Transcripción: dando estructura a datos no estructurados.	9
Campos en el Formulario de Transcripción	10
<b>Datos implícitos: el desafío de transcribir fuentes semiestructuradas.</b>	<b>13</b>
<b>Registros</b>	<b>14</b>
<b>Reflexiones finales</b>	<b>15</b>

# Introducción

Este caso de estudio examina el proceso de transcripción de datos no tabulares o semi estructurados usando DataScribe. Las fuentes históricas pueden venir en un múltiples formas; DataScribe es una herramienta de transcripción que permite a investigadores no sólo transcribir datos, sino también aplicar una estructura tabular a una fuente no estructurada. Este estudio de caso comienza con una descripción general del proyecto *Peste Bubónica en Iquique 1903*, para luego llevar a cabo una caracterización de las fuentes utilizadas. A continuación, el estudio detalla el proceso de organización de las fuentes en ítems de Omeka S para ser usados para crear un proyecto DataScribe en el módulo DataScribe. Posteriormente, el estudio guía el proceso de creación de un formulario de transcripción destinado a capturar información de interés, a propósito de las preguntas históricas planteadas sobre las fuentes. Finalmente, el estudio profundiza en cómo crear los registros en DataScribe y en caminos investigativos que DataScribe permite a los investigadores recorrer.

## Contexto histórico

En mayo de 1903 la peste bubónica arribó a Iquique, ciudad portuaria ubicada en el norte de Chile. La llegada de la peste a Iquique fue parte de una ola de brotes ocurrida en las costas atlánticas y pacíficas de América del Sur, entre finales del siglo XIX y mediados del siglo XX. La peste atacó las costas chilenas entre 1903 y 1928, incluidos el puerto de Valparaíso y la ciudad de Arica.

Para detener la propagación de la peste, el gobierno de Chile conformó una comisión médica destinada a estudiar las causas del brote y establecer medidas de contención. La comisión, formada por un grupo de médicos y bacteriólogos encabezados por el doctor Alejandro del Río, llegó a Iquique el 1 de junio de 1903 y comenzó a trabajar de inmediato en controlar el brote de peste bubónica. A inicios del siglo XX Iquique era de crucial importancia tanto para la industria del salitre como para el gobierno chileno, ya que las exportaciones de salitre aportaban el 25% de los ingresos fiscales anuales del Estado de Chile.

La comisión atendió a los enfermos, estableció un laboratorio bacteriológico y llevó registros detallados de cada paciente contagiado. El doctor Alejandro del Río y las autoridades locales dispusieron un lazareto para atender a los infectos. Luego de tres meses de arduo

trabajo, la comisión y las autoridades locales lograron controlar el brote de peste casi en su totalidad.

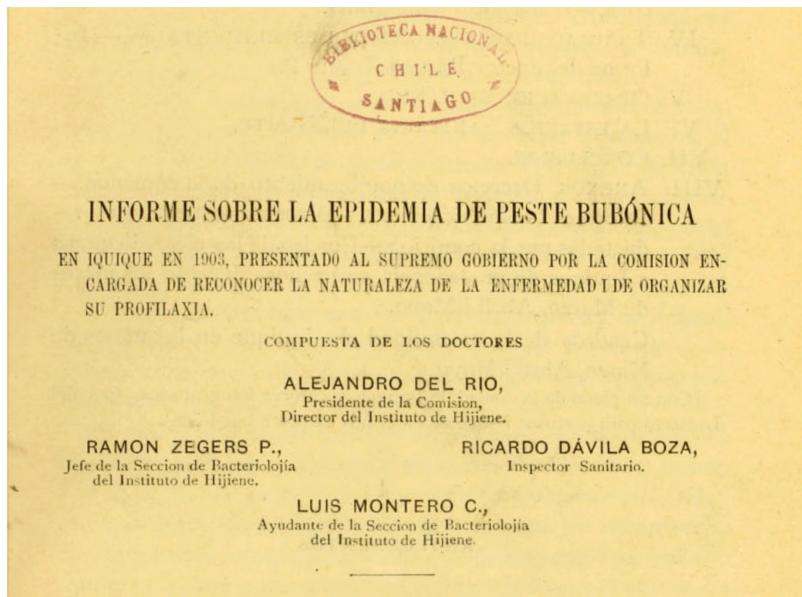


Figura 1. Portada del *Informe Sobre la Epidemia de Peste Bubónica en Iquique en 1903*

A partir de su trabajo, la comisión elaboró varios documentos con información invaluable acerca de las condiciones de vida en Iquique, las actitudes de la gente ante la peste y la mortalidad, la historia clínica de los enfermos, los tratamientos administrados y las medidas higiénicas implementadas por las autoridades. Para este estudio de caso se utiliza el *Informe Sobre la Epidemia de Peste Bubónica en Iquique en 1903* presentado al gobierno de Chile por la comisión médica en 1904. El informe médico contiene información en formato de texto sobre los primeros 167 casos de peste tratados por la comisión médica entre mayo y septiembre de 1903. Para la mayoría de los casos, la lista proporciona dirección, fecha de diagnóstico del caso e inicio de síntomas, y datos sociodemográficos. La naturaleza de la información será analizada con mayor profundidad en la sección acerca de los datos de este estudio de caso.

El proyecto *Peste Bubónica en Iquique 1903* tiene como objetivo examinar el avance de la peste en el puerto de Iquique a partir de la aplicación de metodologías de investigación como *spatial-analysis* y *time-series analysis*. DataScribe es la herramienta que hace posible la transcripción de la información del informe a un formulario tabular, estructurado en filas y columnas. La organización en filas y columnas nos permitirá exportar los datos en un formato adecuado para el análisis computacional.

## Acerca de los datos

Usando DataScribe, este estudio de caso transcribe una lista con los primeros 167 casos tratados por la comisión incluidos en el informe médico. Cada caso en la lista incluye al menos cuatro datos básicos: el número de caso o ID de la persona enferma, el estado del caso (si la persona fue encontrada viva o muerta), la dirección de la vivienda del paciente y la fecha del diagnóstico. La mayoría de los casos también muestran el género, la edad, el nombre del paciente, la fecha en que la persona ingresó al lazareto público (que suele coincidir con el diagnóstico) y la fase del brote en que cada individuo se contagió con la peste. Un subconjunto de casos ofrece información sobre los vínculos entre los casos, sus síntomas y la fecha de aparición de los mismos, y la fecha de muerte o recuperación de los enfermos.

En la calle de Barros Arana 120, manzana 104, foco VI, se recoje el cadáver de Juan Valdes, de 33 años; obs. 56.<sup>a</sup>  
Procedente del hospital ingresa al Lazareto Valentín Villarreal, de 18 años; obs. 57.<sup>a</sup>  
Día 9.—Procedente de la calle de la Estacion 20, manzana 99, foco X, ingresa al Lazareto Manuel Andres Fernandez, niño; obs. 58.<sup>a</sup>  
Día 10.—En un foco nuevo, foco XI, calle de Valparaiso 9, manzana 141-a, se recogen un cadáver i dos enfermos: los hermanos Juan N., Juan A. i José E. Ganga, niños; observaciones 59.<sup>a</sup>, 60.<sup>a</sup> i 61.<sup>a</sup>  
Día 11.—Ingresa al Lazareto la joven Mercedes Diaz, procedente de la calle de la Estacion 18, manzana 99, foco X; obs. 62.<sup>a</sup>  
Día 12.—En la calle de Ramirez 94, manzana 110, foco IX, se recoje enferma i se hace ingresar al Lazareto a la joven Rosa Diaz i se recoje el cadáver de un hombre adulto; observaciones 63.<sup>a</sup> i 64.<sup>a</sup>  
Ingresa al Lazareto el niño Alejandro Villegas, procedente de la calle de Patricio Lynch 154, manzana 53, en la proximidad del foco IX; obs. 65.<sup>a</sup>

Figura 2. Organización textual de la lista de casos del Informe

Es importante señalar que la lista de 167 casos no aparece en una sola sección del informe médico. La lista completa se divide en tres partes, que se pueden encontrar en diferentes secciones del informe. A pesar de estar fraccionada, la lista sigue un formato consistente a lo largo de las secciones; los números de casos se organizan consecutivamente independientemente de las divisiones, y las características básicas de cada caso (identificación, estado, dirección y fecha de diagnóstico) se mantienen en todas las particiones.

Por ejemplo, los casos 1 a 16 se encuentran en las páginas 10 a 15 del informe original y los casos 17 a 75 se encuentran en las páginas 20 a 24.

## Configuración del proyecto DataScribe parte 1: ítems en la instalación de Omeka S

### *Agregar los ítems en Omeka S*

Antes de configurar el proyecto en DataScribe, todos los ítems sujetos a transcripción deben almacenarse en una instalación de Omeka S. Es importante tener en cuenta que DataScribe es un módulo de Omeka S; todos los ítems deben guardarse en Omeka para que la herramienta DataScribe pueda acceder a ellos. Omeka S es la bóveda desde la cual DataScribe extrae ítems para crear conjuntos de datos en un proyecto. Como sistema de gestión de datos, Omeka S permite al usuario gestionar los datos cuidadosamente, no sólo en términos su colocación y almacenamiento, sino también a nivel de información contenida en los metadatos.

Es importante pensar detenidamente cómo las fuentes históricas serán transformadas y almacenadas como ítems de la instalación de Omeka S. En este caso, cada página que contiene secciones de la lista de casos de peste se convertirá en un elemento de la instalación de Omeka S. Esto brinda suficiente flexibilidad para agrupar los ítems en varias formas diferentes, según los tipos de preguntas planteadas sobre los datos y las formas en que deseamos configurar el flujo de trabajo de transcripción. Cada elemento tendrá un archivo multimedia que corresponde a una captura de pantalla en formato png de cada página de la lista con los 167 casos.

Items					Add new item
<input type="checkbox"/>	 Report Peste Bubonica en Iquique, cases 17 to 75, part 4	  	document part	Hernan Adasme Herrera	May 27, 2022
<input type="checkbox"/>	 Report Peste Bubonica en Iquique, cases 17 to 75, part 3	  	document part	Hernan Adasme Herrera	May 27, 2022
<input type="checkbox"/>	 Report Peste Bubonica en Iquique, cases 17 to 75, part 2	  	document part	Hernan Adasme Herrera	May 27, 2022
<input type="checkbox"/>	 Report Peste Bubonica en Iquique, cases 17 to 75, part 1	  	document part	Hernan Adasme Herrera	May 27, 2022
<input type="checkbox"/>	 Report Peste Bubonica en Iquique, cases 17 to 75, part 0	  	document part	Hernan Adasme Herrera	May 27, 2022
<input type="checkbox"/>	 Report Peste Bubonica en Iquique, cases 1 to 16, part 2	  	document part	Hernan Adasme Herrera	May 27, 2022
<input type="checkbox"/>	 Report Peste Bubonica en Iquique, cases 1 to 16, part 1	  	document part	Hernan Adasme Herrera	May 27, 2022

Figura 3. Ítems en Omeka S

Dado que este proyecto sólo contiene 18 ítems, fue posible crear manualmente cada elemento en Omeka S con sus metadatos correspondientes. En la instalación de Omeka S, ha sido creada una plantilla de metadatos llamada Peste en Iquique. Adicionalmente, ha sido seleccionada la clase *documento*, que está disponible en el menú desplegable de clases destinadas a construir los metadatos para los ítems. Dado que *Peste Bubónica en Iquique 1903* es un proyecto pequeño, el proceso de importación de ítems en Omeka S pudo ser realizado manualmente. Para proyectos de mayor envergadura recomendamos usar el complemento CSV Import disponible en la instalación de Omeka S. CVS import es la mejor manera de agregar una gran cantidad de ítems y sus metadatos en Omeka S.

#### *Creación de conjuntos de ítems*

Una vez que todos sus ítems han sido cargados en la instalación de Omeka S, el siguiente paso consiste en crear conjuntos de ítems de Omeka S. DataScribe utiliza conjuntos de ítems de Omeka S como base para los conjuntos de datos de DataScribe. Los conjuntos de ítems Omeka S han sido creados siguiendo las particiones originales de la lista en el Informe. El primer conjunto de ítems de Omeka S titulado *Plague Report. Iquique Chile. Cases 1 a 16*, contiene los cuatro ítems correspondientes a las páginas 10-15 del Informe Médico. En total, se han creado tres conjuntos de ítems diferentes con los ítems que contienen los ítems 10-16, 17-75, 76-167, respectivamente. Decidí agrupar los elementos en correspondencia con las páginas del informe para mantener la coherencia entre el conjunto de elementos y las particiones de la lista del documento original. Si las particiones de varias páginas se hubieran convertido en un solo elemento de varias imágenes, el proyecto DataScribe se habría basado solo en un conjunto de elementos para crear conjuntos de datos. Esa decisión habría reducido severamente las alternativas para organizar el flujo de trabajo de transcripción. Además, dado que la cantidad de información en las particiones de la lista varía, conjuntos de ítems en congruencia con las particiones de la lista captan de mejor manera la naturaleza de la fuente. Esto responde al hecho que existen diferencias entre los detalles incluidos en cada una de las particiones del informe, siendo la más rica en datos la porción que incluye los casos 17-75. La creación de conjuntos de elementos y conjuntos de datos DataScribe en correspondencia con las particiones de página en el informe tiene como objetivo preservar las características del documento original.

Item sets				Add new item set
<input type="checkbox"/>	 Plague Report. Iquique Chile. Cases 17 to 75.	  	Hernan Adasme Herrera	May 27, 2022
<input type="checkbox"/>	 Plague Report. Iquique Chile. Cases 1 to 16.	  	Hernan Adasme Herrera	May 27, 2022
<input type="checkbox"/>	 Plague Report. Iquique Chile. Cases 76 to 167.	  	Hernan Adasme Herrera	May 23, 2022

Figura 4. Conjuntos de datos en Omeka S

## Configuración del proyecto DataScribe, parte 2: creación de un proyecto DataScribe

### *Creando una Peste en Iquique 1903 Proyecto DataScribe*

Una vez configurados y agrupados todos los ítem y conjuntos de ítems en la instalación de Omeka S, todos los elementos necesarios para crear el proyecto Plague in Iquique 1903 DataScribe están disponibles. Para obtener una descripción detallada sobre cómo crear un proyecto de DataScribe, consulte la [documentación de DataScribe](#). El proyecto *Peste Bubónica en Iquique 1903* incluye tres conjuntos de datos; cada uno de esos tres conjuntos de datos corresponde a uno de los conjuntos de ítems en la instalación de Omeka S. La siguiente imagen muestra tres conjuntos de datos: Casos 1 a 16, Casos 17 a 75 y Casos 76 a 167.

Plague in Iquique 1903 (Owner/Reviewer)	
<b>Cases 76 to 167</b>	
Unlocked items	New   In progress
My items (locked to me)	New   In progress   Not approved
Items needing review	All   Needs my re-review
Records needing review	All
<b>Cases 17 to 75</b>	
Unlocked items	New   In progress
My items (locked to me)	New   In progress   Not approved
Items needing review	All   Needs my re-review
Records needing review	All
<b>Cases 1 to 16</b>	
Unlocked items	New   In progress
My items (locked to me)	New   In progress   Not approved
Items needing review	All   Needs my re-review
Records needing review	All

Figura 5. Conjuntos de Datos en el proyecto *Peste Bubónica en Iquique 1903*

Cada conjunto de datos contiene tantos elementos como páginas del informe dedicadas a cada partición de la lista. Por ejemplo, el conjunto de datos Casos 76 a 167 incluye 9 elementos diferentes: Informe Peste Bubónica en Iquique, casos 76 a 167, parte 0, hasta la parte 8. La imagen a continuación muestra los elementos contenidos en el conjunto de datos.

DATASCIENCE: DATASET Cases 76 to 167 · Items				
<a href="#">Dashboard</a> ▶ <a href="#">Projects</a> ▶ <a href="#">Datasets</a> ▶ <a href="#">Items</a>				
1 of 1		1–9 of 9	Prioritized	Descending
<a href="#"></a> <a href="#">Advanced search</a>		No filter	Filter	
Batch actions	Item #	Item	Last submitted	Last reviewed
				Review status
<input type="checkbox"/>	132	Report Peste Bubonica en Iquique, cases 76 to 167, part 1	... Jun 10, 2022	Needs review
<input type="checkbox"/>	133	Report Peste Bubonica en Iquique, cases 76 to 167, part 0	...	New
<input type="checkbox"/>	134	Report Peste Bubonica en Iquique, cases 76 to 167, part 2	...	New
<input type="checkbox"/>	135	Report Peste Bubonica en Iquique, cases 76 to 167, part 3	...	New
<input type="checkbox"/>	136	Report Peste Bubonica en Iquique, cases 76 to 167, part 4	...	New
<input type="checkbox"/>	137	Report Peste Bubonica en Iquique, cases 76 to 167, part 5	...	New
<input type="checkbox"/>	138	Report Peste Bubonica en Iquique, cases 76 to 167, part 6	...	New
<input type="checkbox"/>	139	Report Peste Bubonica en Iquique, cases 76 to 167, part 7	...	New
<input type="checkbox"/>	140	Report Peste Bubonica en Iquique, cases 76 to 167, part 8	...	New

Figura 6. Ítems en los conjuntos de datos

### *Construyendo el Formulario de Transcripción: dando estructura a datos no estructurados.*

La creación del formulario de transcripción es una parte fundamental de la configuración de un proyecto de DataScribe. Para el proyecto *Peste Bubónica en Iquique 1903*, el formulario de transcripción proporciona la base para transformar datos semiestructurados en un conjunto de datos tabulados. Es importante notar que es el formulario de transcripción la herramienta que hace posible la transcripción. DataScribe permite explotar tanto la información cualitativa como cuantitativa disponible en el informe médico de la comisión, a través del enmarque de los datos textuales en un formulario de transcripción transmutable en filas y columnas. Aunque una hoja de cálculo cumple un propósito similar --es decir, organizar y almacenar la información en filas y columnas-- DataScribe garantiza un flujo de trabajo de transcripción consistente y estandarizado, especialmente cuando más de un transcriptor participa en el proceso.

El proceso de transcripción de datos de la lista de contagiados comienza con la evaluación visual de las fuentes para establecer qué campos y tipos de datos son adecuados para capturar con precisión la información disponible. La imagen a continuación muestra un mapa de la información contenida en un caso estándar de la lista de contagiados. La imagen destaca los datos disponibles en el texto que enumera los casos de peste. Es crucial tener en cuenta que se pueden crear tantos formularios de transcripción como conjuntos de datos existan dentro de un proyecto; esto significa que DataScribe permite capturar diferentes tomas de información de la misma fuente. La siguiente imagen destaca los datos explícitos disponibles para la gran mayoría de los casos enumerados, los que son susceptibles de ser transcritos en el proyecto DataScribe.

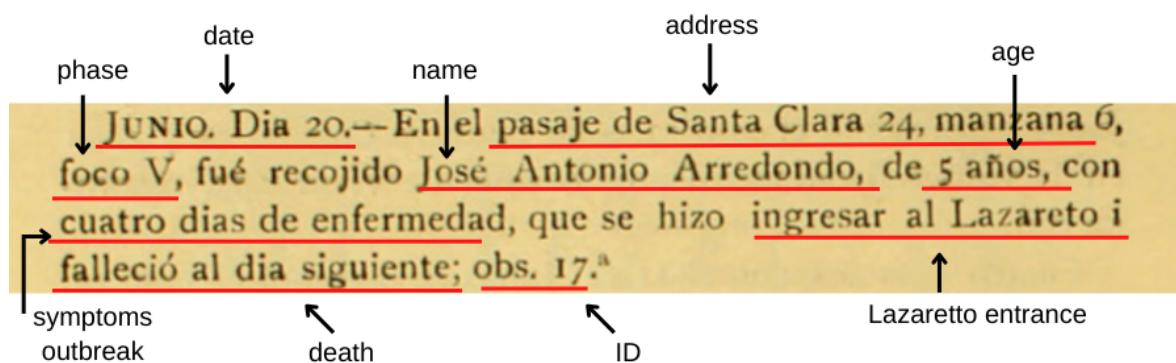


Figura 7. Mapa de los datos incluidos en un caso de la lista

## Campos en el Formulario de Transcripción

El formulario de transcripción contiene 18 campos. Cuatro campos en el formulario de transcripción se definen como campos obligatorios: estos son *case\_ID*, *address*, *diagnosis date*, y *status*. La razón de que los campos *first\_name* y *last\_name* no han sido establecidos como campos obligatorios se debe a que la comisión no pudo identificar un número pequeño de cadáveres y, por lo tanto, sus nombres no aparecen en el informe. Estos cuatro campos requeridos se encuentran en el centro del proyecto de investigación; constituyen el mínimo necesario para identificar cada caso, realizar métodos de investigación de análisis espacial y examinar los datos desde una perspectiva de series temporales.

DATASCIENCE DATASET Cases 1 to 16 · Edit

General Form Builder

Field	Type	Required
case_ID	Number	*
address	Text	*
first name	Text	
last name	Text	
status	Select	*
case_related	Number	
relationship	Text	
gender	Select	
age	Number	
person_type	Select	
job	Text	
symptoms_outbreak	Date	
diagnosis_date	DateTime	
lazaretto_entrance	Date	
death	Date	
recovery	Date	
doctor_name	Select	
phase	Text	

Figura 8. Campos en la ficha de transcripción. Los asteriscos indican los campos requeridos.

Seis campos en la ficha de transcripción han sido establecidos como datos de texto: *first\_name*, *last\_name*, *address*, *relationship*, *phase* y *job*. Dentro de esta categoría, tanto la fase como las relaciones ofrecen etiquetas de entrada para ayudar en el proceso de transcripción. Aunque la fase de campo se refiere conceptualmente al número consecutivo de brotes de peste, el campo se estableció como un texto porque el informe original identifica las fases del brote con números romanos.

**relationship**  
relationship of the case related with the main case

enter the relationship

Is missing  Is illegible

Reset value

**gender**  
gender of the case

Select female

sister  
brother  
mother

Figura 9. Etiquetas de entrada en el campo que captura las relaciones entre casos

Tres campos en el formulario son numéricos: *case\_ID*, *case\_related* y *age*. El campo *case\_ID* captura la identificación original utilizada por la comisión en el informe; este campo se ha establecido como el campo principal, es decir, el campo por el cual se identifica el registro. El campo *case\_related* hace referencia a la identificación de otro caso en la lista que está relacionado con el caso principal que se transcribe. Está diseñado conceptualmente para ser un campo externo autorreferencial o recursivo ya que es un campo que hace referencia a la misma tabla.

Cinco campos en el formulario de transcripción corresponden al tipo de datos de fecha y fecha y hora. La transcripción recopila la fecha de diagnóstico (campo *diagnosis\_date*, que también es un campo obligatorio), las fechas de surgimiento de los síntomas (campo *symptoms\_outbreak*) y de entrada al lazareto (campo *lazzaretto\_entrance*). El formulario de transcripción captura además de fechas de muerte y recuperación (campos *death* y *recovery*, respectivamente). El campo *diagnosis\_date* es un tipo de datos de fecha y hora, que permite al usuario ingresar la hora precisa, disponible para algunos casos. Tanto la fecha como la fecha y la hora se exportarán en formato ISO 8601 mediante el calendario gregoriano. Estos datos de fecha y hora son cruciales para realizar análisis de series temporales computacionales y de esta forma rastrear el avance del brote de peste.

Finalmente, cuatro campos son de tipo seleccionar (select) y crean un menú desplegable con opciones para elegir: *status*, *gender*, *person\_type*, y *doctor\_name*. El campo *status* ofrece dos opciones: vivo o muerto desde un menú desplegable. Como se explicó anteriormente, este es un dato que debe inferirse, pues se menciona exclusivamente cuando la comisión médica encuentra y diagnostica un cadáver. El campo *gender* muestra un menú

desplegable con solo dos alternativas: masculino y femenino. Para la mayoría de los casos, el género de la persona enferma debe inferirse del nombre y el uso de género del idioma español. El campo *person\_type* se utiliza cuando no se menciona explícitamente la edad del paciente, pero la fuente describe a la persona enferma como un bebé, un niño o un adulto. Finalmente, el campo *doctor\_name* captura el nombre del médico que realizó el diagnóstico y los primeros tratamientos, en caso de ser mencionado por la fuente.

The screenshot shows a transcription interface. On the left, there is a preview of a document page titled "1. plague\_cases\_17\_75\_0.png". The document contains historical text about plague cases in Madrid. On the right, there is a form with the following fields:

- case\_ID \***: A field with the placeholder "ID case according to document". It includes a "Number" input field, a "Reset value" button, and checkboxes for "Is missing" and "Is illegible".
- address \***: A field with the placeholder "address of the case". It includes a "Text" input field, a "Reset value" button, and checkboxes for "Is missing" and "Is illegible".
- first name**: A field with the placeholder "first name of the sick person". It includes a "enter first name" input field, a "Reset value" button, and checkboxes for "Is missing" and "Is illegible".
- last name**: A field with the placeholder "last name of the sick person". It includes a "Reset value" button and checkboxes for "Is missing" and "Is illegible".

Figura 10. Área de trabajo de transcripción

Los mismos campos descritos pueden ser agrupados según tres familias de datos: datos espaciales, datos temporales y datos relacionales. La dirección de cada caso de peste se registra como un texto que incluye la dirección y el número de bloque. Los campos que capturan varias fechas que se encuentran en la fuente histórica traducen la información a un formato de tipo de datos estandarizado que, a su vez, garantiza la homogeneidad entre los formatos de fecha. Finalmente, algunos casos brindan información sobre cómo se relacionan los casos entre sí; es decir, qué casos de peste corresponden a un cúmulo de casos entre personas emparentadas. Los campos *case\_related* y *relation* han sido diseñados para obtener el ID del caso como campo externo, y una característica del tipo de relación entre casos.

# Datos implícitos: el desafío de transcribir fuentes semiestructuradas.

Es importante tener en cuenta que algunos fragmentos de información no son explícitos y deben inferirse de la fuente. Por ejemplo, el campo estado (*status*), que se refiere a la condición del caso al momento del diagnóstico --si el enfermo fue encontrado vivo o muerto-- no se menciona explícitamente en el listado; únicamente los casos fallecidos al momento de diagnóstico fueron descritos en el informe como “encontrados muertos” o “cadáver encontrado en”. Sin embargo, aunque esta información no se menciona de forma explícita, es posible inferir que los casos que no están etiquetados como muertos fueron encontrados vivos por los doctores. En consecuencia, el campo *status* en los formularios de transcripción permite al transcriptor seleccionar la opción “vivo” o “muerto” de un menú desplegable. Lo mismo ocurre con el género; como se mencionó antes, para la mayoría de los casos, el género del paciente debe inferirse del género usado en la gramática española. Finalmente, ambos campos *diagnostic\_date* y *lazarreto\_entrance* coinciden temporalmente, y corresponden a la fecha general que organiza cómo se presentan los casos en el informe. Es importante tener en cuenta que la lista está organizada por fecha; la fecha precisa en que los casos fueron diagnosticados por los médicos es el criterio principal con que los autores del informe ordenaron los casos y determinaron la ID de la observación. Dicho de otra manera, es la fecha de diagnóstico la que determina el orden de la lista, y no el inicio de los síntomas de los pacientes.

Para la transcripción de información difícil de transcribir es importante agregar directrices a cada conjunto de datos en un proyecto de DataScribe. La transcripción de fuentes semiestructuradas plantea desafíos debido a que la información textual no siempre sigue un formato estándar: la estructura del texto cambia de un caso a otro. Las directrices permiten agregar instrucciones específicas e información útil para garantizar la coherencia del trabajo de transcripción durante todo el proceso de transcripción.

# Registros

La siguiente imagen muestra una captura de pantalla de los registros en su forma final. El número de registro 62 en la imagen corresponde al mismo dato que se muestra en la imagen del mapa de datos. La imagen muestra cómo los campos, *case\_id*, *address*, *first name*, *last name*, *gender*, *age*, *person\_type*, *phase* y *symptoms\_outbreak* se han traducido de datos textuales semiestructurados a información almacenada en filas y columnas.

The screenshot shows a data records interface with the following details:

Record #	Created by	Created	case_ID	address	first name	last name	status	case_related	relationship	gender	age	person_type	job	phase	symptoms_outbreak
62	Hernan Adasme Herrera	Jul 14, 2022	17	Pasaje Santa Clara 24, manzana 6	Jose Antonio	Arredondo	active case	null	null	male	5	kid	null	5	1903-06-16
63	Hernan Adasme Herrera	Jul 14, 2022	18	Vivar 289, manzana 196	Miguel	Angulo	active case	null	null	male	19	adult	null	2	1903-06-14
64	Hernan Adasme Herrera	Jul 14, 2022	19	Zegers 85, manzana 175	Tomas	Briones	dead	null	null	male	10	kid	null	2	1903

Figura 11. Registros en el conjunto de datos

## Reflexiones finales

Aunque DataScribe ha sido diseñado para ayudar en la transcripción de datos históricos estructurados, el software es lo suficientemente flexible como para hacer posible la transcripción de fuentes históricas textuales y semiestructuradas a un formato tabulado. DataScribe proporciona las herramientas para crear la arquitectura que enmarcar las piezas de información en filas y columnas. Para este proyecto, usamos DataScribe para construir una estructura externa que disecciona la fuente histórica, captura información de interés y reorganiza la fuente de una manera que permite aplicar métodos computacionales de análisis. DataScribe minimiza los errores de transcripción al hacer que los datos recopilados sean más consistentes y estandarizados, especialmente cuando el flujo de trabajo incluye más de un transcriptor.

La transcripción no estructurada conlleva sus propios desafíos. Dado que estamos traduciendo un formato textual en tablas con filas y columnas, el trabajo de transcripción requiere directrices claras para ayudar a los transcriptores a cumplir con los criterios de transcripción. El flujo de trabajo también requiere una revisión constante, porque el estilo y los detalles contenidos en la información escrita varían entre registros. En el caso de las fuentes utilizadas para *Peste Bubónica en Iquique 1903* la cantidad de información que no es explícita y debe ser inferida hace que el proceso de transcripción sea difícil y propenso a errores. No todos los campos pueden llenarse simplemente trasladando la información de la de una casilla en fuente a la casilla correspondiente en el formulario de DataScribe. La transcripción de esas piezas particulares de información requiere una buena cantidad de *close reading* para crear un registro de transcripción. Los formularios de transcripción de DataScribe se pueden actualizar constantemente para seguir mejorando la calidad y la funcionalidad del trabajo de transcripción. DataScribe permite crear diferentes conjuntos de datos para aplicar varios formularios de transcripción a la misma fuente histórica. Todos los formularios de transcripción ayudarán a capturar material, dejando sin embargo cierta información atrás. DataScribe aspira a preservar aquellos datos en riesgo de quedar en el olvido.