

De la narrativa documental al dato estructurado: anotación semántica con Recogito en un corpus de pintores coloniales en México

From Documentary Narrative to Structured Data: Semantic Annotation with Recogito in a Corpus of Colonial Mexican Painters

Maria Laura FLORES BARBA

mfloresb@uwo.ca

Western University, Canadá

ID <https://orcid.org/0000-0003-3354-8358>

Cita recomendada:

Flores Barba, M. L. (2025). De la narrativa documental al dato estructurado: anotación semántica con Recogito en un corpus de pintores coloniales en México. *Publicaciones de la Asociación Argentina de Humanidades Digitales*, 6, e072. <https://doi.org/10.24215/27187470e072>

RECIBIDO: 12/09/2025 ACEPTADO: 10/12/2025 PUBLICADO: 16/12/2025

RESUMEN

Este artículo presenta un ejercicio metodológico para transformar fuentes históricas narrativas en datos estructurados mediante la anotación semántica con Recogito. En el marco de una investigación doctoral sobre pintores en México (1680–1730), se trabajó con un corpus de 23 documentos parroquiales vinculados a la familia Cuentas, una dinastía activa en varias ciudades novohispanas. Se describe el flujo de trabajo desde la transcripción hasta la exportación de datos, incluyendo el uso de vocabularios personalizados, estrategias de desambiguación y jerarquías de etiquetas. El estudio destaca tanto el potencial de Recogito como herramienta adaptable para generar datasets reutilizables, como los límites inherentes a la fuente seleccionada. Se argumenta que la conversión de lo narrativo a lo tabular es una operación interpretativa que requiere un modelo de datos claro, protocolos consistentes y una lectura crítica de las ausencias documentales.

PALABRAS CLAVE: anotación, modelización, Recogito, limpieza de datos, análisis relacional.

ABSTRACT

This article presents a methodological exercise in transforming narrative historical sources into structured data through semantic annotation with Recogito. Within the framework of a doctoral study on painters in Mexico (1680–1730), a corpus of 23 parish records related to the Cuentas family, a dynasty active in several New Spanish cities, was analyzed. The workflow is detailed from transcription to data export, including the use of controlled vocabularies, disambiguation strategies, and tag hierarchies. The study highlights Recogito's potential as an adaptable tool for generating reusable datasets, as well as the inherent limitations of the selected sources. It argues that converting narrative into tabular form is an interpretive operation requiring a clear data model, consistent protocols, and a critical reading of documentary absences.

KEYWORDS: annotating, modeling, Recogito, cleanup, relational analysis.

1. INTRODUCCIÓN

En el marco de mi tesis doctoral, diseñé y construí una base de datos relacional para estudiar las redes sociales de pintores en México entre 1680 y 1730 (Flores Barba, 2025). La historiografía sobre la pintura colonial mexicana había documentado abundantemente nombres, obras y trayectorias individuales, pero rara vez se había propuesto articular esas trayectorias de manera relacional. Mi objetivo fue doble: por un lado, centralizar y organizar la información dispersa sobre pintores y, por otro, crear una infraestructura de datos que permitiera visualizar vínculos, genealogías y agrupaciones locales o regionales.

Este enfoque implicó un desafío metodológico importante: muchas de las fuentes utilizadas, tanto primarias como secundarias, estaban escritas en forma narrativa o documental y no eran fácilmente transferibles a un modelo tabular. Aun cuando la información fuera explícita, como por ejemplo “Diego de Cuentas, pintor, hijo legítimo de Diego de Cuentas y Teresa Patiño, vecinos de Valladolid”, su traducción a un lenguaje de datos estructurados requería decisiones metodológicas precisas tales como ¿cómo identificar a los actores?, ¿cómo registrar sus relaciones?, ¿cómo codificar atributos como género, ocupación o categoría racial?

Durante la fase de recolección de datos, encontré en Recogito¹ una herramienta particularmente útil para abordar este tipo de problemas. Desarrollado por el proyecto Pelagios Network, Recogito es una plataforma de código abierto que permite anotar documentos con etiquetas semánticas (personas, lugares, eventos) y asociar esas entidades a fuentes controladas como *gazetteers*. Si bien la herramienta fue pensada originalmente para proyectos de anotación colaborativa y georreferenciación, en mi caso la utilicé como un medio para transformar textos históricos en datos estructurados, listos para ser procesados y analizados.

Este artículo presenta un ejercicio metodológico basado en cómo usé Recogito para extraer información estructurada a partir de documentos históricos, adaptando su funcionalidad a las necesidades de mi investigación. No se trata de un tutorial ni de un estudio cuantitativo, sino de una reflexión sobre los usos posibles de una herramienta digital en el marco de un proyecto de historia digital del arte. El caso que analizo, un conjunto de 23 documentos relacionados con la familia Cuentas, una dinastía de pintores activos en Michoacán, Ciudad de México, Aguascalientes y Guadalajara, muestra cómo la anotación

¹ Disponible en: <https://recogito.pelagios.org/>.

semántica puede servir como puente entre el archivo y la base de datos y abrir nuevas posibilidades analíticas sin requerir programación avanzada ni automatización total.

1.1. Contexto: base de datos y fuentes

Mi investigación doctoral se centró en un período de transición dentro del ámbito pictórico novohispano. Se trata de una etapa de renovación y cambio estilístico que comienza a manifestarse en las obras de Cristóbal de Villalpando (c. 1649–1714) y Juan Correa (c. 1645–1716), alcanza un punto de inflexión con los hermanos Juan (1675–1728) y Nicolás Rodríguez Juárez (1667–1734) y culmina con Miguel Cabrera (1695–1768). Esta transformación no fue solamente artística, sino también institucional, influida por las regulaciones gremiales y el contexto intelectual del criollismo. Antonio Rubial caracteriza esta etapa como la “época barroca” (1640–1750), describiéndola como un momento en el que la identidad criolla se cristalizó y los símbolos de lo novohispano, su historia, sus santos y sus imágenes milagrosas, se arraigaron profundamente en la producción artística (Rubial García, 2010). Todo esto convierte al periodo en un campo de estudio especialmente rico y complejo para analizar las redes de pintores que moldearon el paisaje artístico del México colonial.

Para poder analizar estas trayectorias desde una perspectiva relacional, diseñé una base de datos a medida, compuesta por 18 tablas y más de 3000 registros de personas, artefactos, eventos y relaciones. Cada afirmación registrada en la base de datos debía estar respaldada por una fuente concreta, identificada mediante un ID único (Pasin y Bradley, 2015). Esto me permitió no solo organizar la información, sino también trazar su origen, distinguir entre datos factuales e interpretaciones y dar cuenta de contradicciones o disputas historiográficas.

Los sets de datos que usé para alimentar mi base de datos pueden dividirse en cinco categorías, que responden tanto al tipo de documento como a su grado de estructuración, y reflejan las distintas estrategias empleadas para convertirlos en datos reutilizables:

1. Fuente fundacional: El punto de partida para construir el listado inicial de actores fue la “Nómina de pintores coloniales” incluida por Manuel Toussaint en su libro *Pintura colonial en México* (1965). Esta nómina, producto del trabajo de archivo y campo realizado por Toussaint, reúne nombres de pintores, sus lugares de trabajo cuando eran conocidos, algunas fechas clave y referencias. Esta fuente me permitió poblar los primeros registros de actores en la base de datos y establecer un punto de partida para la posterior incorporación de datos más detallados.
2. Fuentes de artefactos: para vincular obras de arte con los actores identificados, recurrió a la Mediateca del INAH.² Esta plataforma digital ofrece acceso a miles de registros de objetos patrimoniales como pinturas, esculturas, documentos y fotografías provenientes de distintas colecciones nacionales (Vázquez, 2019). La Mediateca implementa el protocolo OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting), lo que me permitió cosechar metadatos de manera automatizada. A través de este protocolo descargué descripciones de obras atribuidas a los pintores registrados, autorías, ubicaciones, fechas y referencias institucionales. Estos metadatos fueron luego limpiados, normalizados y adaptados con OpenRefine³ para integrarlos dentro de la base de datos.

² Disponible: en https://mediateca.inah.gob.mx/islandora_74/.

³ OpenRefine es una herramienta de organización, limpieza y transformación de datos, gratuita y de código abierto. Disponible en <https://openrefine.org/>.

3. Monografías sobre pintores específicos: este conjunto incluyó estudios biográficos dedicados a artistas particulares como Juan Correa (Vargas Lugo et al., 1991), Cristóbal de Villalpando (Gutiérrez Haces, 1997), Ignacio Berben (Ramírez, 2009), Pedro López Calderón (Mayorga Chamorro, 2019) y Antonio de Torres (Gila Medina, 2015), entre otros. Se trata de publicaciones académicas en forma de libros, tesis y artículos, que recuperan documentos inéditos, interpretan trayectorias y ofrecen contextos regionales o iconográficos.
4. Catálogos de documentos históricos: esta categoría agrupa las publicaciones del Instituto Nacional de Antropología e Historia (INAH) y el Instituto de Investigaciones Estéticas (IIE) que ofrecen catálogos de documentos localizados en archivos coloniales en México (por ejemplo, González Franco et al., 1994; Ramírez Montes y Luckie, 1993). Aunque no reproducen el documento completo, estos catálogos proporcionan resúmenes descriptivos (en formato de fichas) con fechas, lugares, actores mencionados y referencias archivísticas precisas. Son fuentes semiestructuradas que permiten extraer datos de forma relativamente directa.
5. Fuentes archivísticas directas: finalmente, también recopilé documentación primaria procedente de distintos archivos históricos, tanto físicos como digitales. Consulté materiales del Archivo General de la Nación (AGN) en la Ciudad de México, el Archivo Histórico de Jalisco (AHEJ), el Archivo de la Real Audiencia de Guadalajara y la plataforma FamilySearch, que contiene registros parroquiales digitalizados.

A partir de todas estas fuentes, busqué identificar actores relevantes además de pintores, tales como parientes, testigos, clientes, maestros y aprendices y registrar sus atributos: nombre, apellido, género,

ocupación, nivel profesional, casta (cuando aparecía) y las relaciones que mantenían entre sí (consanguinidad, jerarquía profesional, vínculos de lugar).

Esta tarea resultó compleja. Muchas de las fuentes fueron digitalizadas como imágenes, por lo que requerían un proceso de reconocimiento óptico de caracteres (OCR) y normalización. Una vez convertidas en texto, el desafío era organizarlas en una tabla: extraer los datos pertinentes sin perder el contexto y asegurarse de que cada elemento fuera compatible con la estructura de la base de datos. En ese momento, exploré varias herramientas de anotación y finalmente opté por Recogito. Este artículo presenta la metodología que diseñé para agregar las fuentes documentales a la base de datos.

1.1.1. El corpus: la dinastía Cuentas

El conjunto documental que utilicé en este ejercicio metodológico está centrado en la familia Cuentas, una dinastía de pintores novohispanos cuya historia me permitió estudiar tanto las prácticas del oficio como las relaciones familiares, laborales y sociales en el ámbito artístico de la Nueva Galicia para mi tesis de maestría (Flores Barba, 2014). Este núcleo familiar resulta particularmente interesante porque permite seguir, casi sin interrupciones, el recorrido vital de su fundador, Diego de Cuentas, así como la continuidad de su legado a lo largo de tres generaciones.

Según mi investigación documental, Diego Alencaster de Cuentas⁴ nació en Acámbaro en 1654, hijo natural de Juan Martínez de Araujo, quien fuera abogado de la Real Audiencia y luego presbítero, y de María de Celis y Cuentas. Su nacimiento estuvo marcado por el ocultamiento deliberado de su filiación: su madre mantuvo su embarazo en secreto y lo depositó en la puerta de la casa de su tío José de Cuentas, quien lo crió

⁴ El nombre con el que firmaba sus pinturas y documentos era Diego de Cuentas, pero en algunas fuentes se le nombra Diego Alencaster de Cuentas.

como suyo y le dio su apellido. A los diez años, este tío lo llevó a la Ciudad de México para que aprendiera el oficio de pintor, aunque aún se desconoce con quién se formó. Más tarde regresó a Valladolid (hoy Morelia), donde se casó y tuvo seis hijos, entre ellos Diego y Juan Antonio, quienes también se dedicarían a la pintura.

En 1705, la familia se trasladó a Guadalajara. Allí Diego de Cuentas estableció un obrador multigeneracional, donde trabajaban sus hijos, un nieto, dos aprendices y, eventualmente, también sus yernos (Flores Barba, 2009). En su testamento, fechado en 1744, se presenta como el patriarca de una red compleja de relaciones domésticas y laborales (Orendáin, 1951). Este documento no solo detalla su patrimonio, sino también los vínculos afectivos y jerárquicos que organizaban la vida del taller. Curiosamente, en ese testamento, Diego omite los apellidos de sus padres y utiliza en cambio los de sus abuelos, Juan Martínez de Alencaster y María de Sandoval y Gamboa, posiblemente para respetar la voluntad de su madre de mantener en secreto el escándalo de su nacimiento.

Además de su papel como jefe de taller, Diego de Cuentas fue un pintor reconocido por sus contemporáneos. Participó como perito en al menos un caso famoso: el examen de una imagen supuestamente milagrosa hallada en una mina (Mota Padilla, 1870, p. 443), de la misma manera en que Miguel Cabrera haría lo propio con la imagen de la Virgen de Guadalupe (Cabrera, 1756). Esta participación confirma su estatus como autoridad artística en el occidente novohispano.

Tras su muerte, el taller dejó de figurar en los registros documentales hasta 1757, año de la única pintura firmada por su nieto, Francisco Xavier Cuentas, quien reactivó parcialmente la tradición familiar. Francisco Xavier se casó con la hija de un notario prominente de Guadalajara y enseñó el oficio a sus propios hijos. Uno de ellos, Manuel Cuentas y Santamaría, llegó a ser profesor en la Academia de Dibujo del Real Consulado de Guadalajara (Diego-Fernández Sotelo y Mantilla Trolle,

2003, p. xcvii). Este linaje artístico, que abarca más de medio siglo, permite examinar de forma detallada cómo se heredaban, enseñaban y transformaban las prácticas artísticas en una región que ha sido tradicionalmente considerada periférica frente a la centralidad de la Ciudad de México.

La historia completa de la familia Cuentas combina documentos notariales, registros parroquiales, firmas en obras de arte y testamentos. La riqueza de esta documentación la convierte en un caso paradigmático para probar herramientas digitales como Recogito, que permiten trazar con precisión la densidad de relaciones personales, profesionales y espaciales que estructuran la vida de un obrador colonial.

2. METODOLOGÍA

Recogito es una herramienta de anotación semántica que permite etiquetar entidades como personas, lugares y eventos dentro de un texto, y asociarlas con metadatos. Aunque fue pensada para la georreferenciación de fuentes históricas y la colaboración en proyectos de historia antigua y medieval, su diseño abierto y flexible la hace útil también en contextos modernos o coloniales, y en proyectos individuales.

Lo que más me interesó de Recogito fue la posibilidad de anotar textos sin necesidad de programación, exportar los resultados en formato CSV para posteriormente poder importarlos a mi base de datos, usar etiquetas personalizadas para añadir atributos y asociar comentarios libres a cada entidad. En vez de utilizar Recogito para georreferenciar o colaborar, lo adapté a mis propios fines: crear un flujo de trabajo que me permitiera pasar de la transcripción paleográfica a una tabla estructurada, manteniendo el vínculo con la fuente original.

2.1. Definición previa del modelo de datos

Antes de comenzar a anotar en Recogito, elaboré un esquema de extracción basado en los campos de mi base de datos. Estos incluían:

- ID de la fuente: código único para cada documento. En el caso de las fuentes primarias citadas en una fuente secundaria, registré ambas para no perder la cadena de referencias.
- Lugar del evento: ciudad o villa donde ocurrió un bautismo, matrimonio, examen o testamento. Cuando se mencionaban lugares específicos como una iglesia o un convento, registré dichos detalles.
- Nombre del actor: dividido en nombre(s) y apellido(s) y separados en campos distintos.
- Género: dividido en cinco categorías – masculino, femenino, grupo mixto, grupo exclusivamente femenino, grupo exclusivamente masculino. Esta categorización me permitió incluir en la base de datos entidades grupales como órdenes religiosas, cofradías, familias, etc.
- Ocupación u oficio: registrada tal como aparece en la fuente (pintor, soldado, escribano, etc.) y basada en el *Tesoro de Arte y Arquitectura del Getty* (AAT)⁵.
- Nivel de ocupación: aprendiz, maestro, doctor, licenciado, etc.
- Casta o categoría racial: cuando figura en la fuente y transcrita tal cual aparece registrada (mestizo, mulato, español, etc.). Este dato resultó sumamente interesante cuando hice el análisis de redes, sobre todo porque durante la época estudiada la reglamentación gremial prohibía a personas no españolas tener su propio taller. Es decir, miembros de otras castas podían trabajar únicamente como aprendices y oficiales dentro de talleres liderados por pintores

⁵ Disponible en: <https://www.getty.edu/research/tools/vocabularies/aat/index.html>.

españoles y/o criollos. Sin embargo, al ser la casta una categoría subjetiva y sumamente flexible, estas regulaciones no funcionaron en la práctica.

- Relaciones: solamente registré las descendentes (padre/madre de, maestro de) o simétricas (hermano de, colega de), así como relaciones filiales de sangre, a excepción de padrino/madrina y compadre/comadre, ya que el compadrazgo fue uno de los mecanismos a los que la sociedad novohispana recurrió para ascender en la escala social y diversificar sus redes.

2.2. Optimización del proceso de anotación: iteraciones, desambiguación y vocabularios personalizados

Aunque Recogito ofrece una interfaz accesible y flexible para la anotación de entidades, alcanzar una estructura de datos exportable sin errores requería una estrategia cuidadosamente diseñada. El trabajo que presento aquí no fue lineal; desarrollé el procedimiento a partir de pruebas e iteraciones sucesivas que me permitieron refinar tanto la estructura del esquema de etiquetas como el flujo de trabajo general.

Una de las primeras decisiones clave que tomé fue consolidar los 23 documentos del corpus en un único archivo de texto. Esto facilitó una de las funcionalidades más potentes de Recogito, la desambiguación. Al trabajar todo el corpus en conjunto y comparar los documentos entre sí, pude enlazar cada mención de una persona al mismo identificador, sin necesidad de esperar hasta la etapa de limpieza en OpenRefine. Aunque Recogito ofrece la opción de reconocimiento automático de entidades (NER por sus siglas en inglés)⁶ para identificar nombres propios de

⁶ El reconocimiento automático de entidades nombradas (NER) es una técnica de procesamiento de lenguaje natural que utiliza modelos de inteligencia artificial para identificar y clasificar menciones específicas en un texto desestructurado o semiestructurado, como nombres de personas, lugares, fechas, organizaciones o conceptos relevantes.

personas y de lugares, no usé esta función por varias razones: en primer lugar, la ortografía de los nombres en los documentos históricos no es homogénea, por lo que el algoritmo reconoce como dos personas distintas a Diego Quentas y Diego Cuentas; en segundo lugar, hay demasiados homónimos, por lo que identifica a Diego de Cuentas padre como la misma persona que Diego de Cuentas hijo; finalmente, mi método requería que solo se mencionara la misma persona una sola vez por documento para facilitar la limpieza posterior de los datos.

Esta capacidad de desambiguar directamente en Recogito me permitió establecer con mayor claridad quién era quién en cada fuente, reducir ambigüedades y evitar duplicaciones. Para facilitar aún más este proceso, construí un archivo auxiliar que cargué como vocabulario de etiquetado (Tagging Vocabulary) en Recogito. Este archivo contenía, en formato CSV, una lista de todas las personas ya registradas en mi base de datos, sus nombres y sus URIs internas, así como los oficios incluidos en mi taxonomía ocupacional.

La estructura de este archivo seguía el formato requerido por Recogito: "etiqueta, URI". Aunque la URI es técnicamente opcional, descubrí que incluirla desde el principio facilitaba el trabajo de desambiguación y garantizaba que las entidades anotadas pudieran ser identificadas únicamente al exportar los archivos CSV e importarlos en mi base de datos.

Isabel Benzor Meza	https://people-things-traces.org/actor/02832
Nicolás de Cuevas	https://people-things-traces.org/actor/02833
Antonia Flores de Villarreal	https://people-things-traces.org/actor/02834
Domingo Sainz de Leyva	https://people-things-traces.org/actor/02835
Domingo Antonio Sainz de Leyva	https://people-things-traces.org/actor/02836
María de Mendoza	https://people-things-traces.org/actor/02837
Juan de Orozco	https://people-things-traces.org/actor/02838
Juan de Garay	https://people-things-traces.org/actor/02839
Pedro José Mariano Berbén McDowell	https://people-things-traces.org/actor/02840
occupation_name	occupation_uri
painters (artists)	http://vocab.getty.edu/aat/300025136
illuminators (manuscript artists)	http://vocab.getty.edu/aat/300025122
friars	http://vocab.getty.edu/aat/300266789
healers	http://vocab.getty.edu/aat/300259528
clergy	http://vocab.getty.edu/aat/300025747
appraisers	http://vocab.getty.edu/aat/300025829
deputies	http://vocab.getty.edu/aat/300250472
bishops (prelates)	http://vocab.getty.edu/aat/300025756
merchants	http://vocab.getty.edu/aat/300025245
captain general	https://www.wikidata.org/entity/Q107147053

Figura 1. Detalle de la lista de etiquetas para cargar en Recogito. Esta lista muestra las últimas etiquetas correspondientes a *actores* y las primeras a *ocupaciones*. En las URIs se puede ver que los registros de los actores son URIs internas que corresponden a mi proyecto que eventualmente publicaré bajo el nombre *people-things-traces* y las de ocupaciones pertenecen al vocabulario del Getty, a excepción de la última, que tomé de Wikidata. elaboración propia.

Además del vocabulario de personas y ocupaciones, definí un proceso estandarizado para la secuencia y el contenido de las etiquetas aplicadas a cada entidad. Esto fue crucial para asegurar que, al exportar los datos, todas las columnas generadas mantuvieran una lógica estructural constante. Las etiquetas fueron ordenadas jerárquicamente según su importancia y su frecuencia de aparición. Por ejemplo:

1. ID de la fuente: siempre presente, primera etiqueta para todas las entidades. Me permitió controlar visualmente (usando la funcionalidad de coloración por etiquetas de Recogito) si todas las entidades de un documento habían sido anotadas correctamente.

2. ID del actor: si la persona ya existía en la base de datos, agregué su identificador como segunda etiqueta. Para ello, usé la lista de etiquetas cargada en Recogito. Si el actor no aparecía en la lista, apliqué la etiqueta "nueva persona".

3. Género, casta, ocupación, nivel profesional, etc.: estas etiquetas las agregué aunque no siempre hubiera información disponible. Cuando faltaba el dato, conservé el nombre de la etiqueta (por ejemplo, "casta"), a manera de *placeholder* o comodín para que la estructura del CSV no se alterara.

Este último punto es especialmente importante. Si una entidad no tiene una de las etiquetas y simplemente se omite, al exportar el CSV y procesarlo en OpenRefine, el número de columnas generadas no es equivalente. Esto complica enormemente el proceso de limpieza, ya que valores de un mismo tipo (por ejemplo, casta) pueden aparecer en columnas diferentes según el número de etiquetas presentes en cada fila, como muestra la Figura 3. Al mantener el mismo número y orden de etiquetas en todas las entidades, se garantiza que cada tipo de información esté siempre en la misma columna y pueda integrarse a la base de datos sin necesidad de un proceso adicional de limpieza y organización.

QUOTE_TRANSSCRIPTION	TYPE	TAGS	COMMENTS
Petrona Basurto	PERSON	Cuentas 15	
Manuel de Meza Cervantes	PERSON	Cuentas 15 masculino	Manuel de, Meza Cervantes
Antonio de Cuentas	PERSON	Cuentas 16 masculino español	Juan Antonio de, Cuentas
Agustina De la Torre	PERSON	Cuentas 16 femenino española	Agustina de la, Torre
Manuel Nieto	PERSON	Cuentas 16 masculino	Manuel, Nieto
Eusebio Flores	PERSON	Cuentas 16 masculino	Eusebio, Flores
Vicente Preciado de Elizalde	PERSON	Cuentas 16 masculino bachiller	Vicente, Preciado de Elizalde
Pedro Padilla	PERSON	Cuentas 17 masculino Sacerdotes provisor vicario doctor	Pedro, Padilla
Juan Manuel de los Reyes	PERSON	Cuentas 17 masculino español	Juan Manuel de los, Reyes Vecino de Guadalajara
Maria de los Reyes	PERSON	Cuentas 17 femenino	Maria de los, Reyes

Figura 2. Tabla en OpenRefine que muestra los resultados del archivo exportado de Recogito. En la columna *Tags* se recogen los datos anotados como etiquetas y en la columna *Comments* los anotados como comentarios. Fuente: elaboración propia.

Petrona Basurto	PERSON	Cuentas 15								
Manuel de Meza Cervantes	PERSON	Cuentas 15	masculino						Manuel de, Meza Cervantes	
Antonio de Cuentas	PERSON	Cuentas 16	masculino	español					Juan Antonio de, Cuentas	
Agustina De la Torre	PERSON	Cuentas 16	femenino	española					Agustina de la, Torre	
Manuel Nieto	PERSON	Cuentas 16	masculino						Manuel, Nieto	
Eusebio Flores	PERSON	Cuentas 16	masculino						Eusebio, Flores	
Vicente Preciado de Elizalde	PERSON	Cuentas 16	masculino	bachiller					Vicente, Preciado de Elizalde	
Pedro Padilla	PERSON	Cuentas 17	masculino	Sacerdotes	provisor	vicario	doctor		Pedro, Padilla	
Juan Manuel de los Reyes	PERSON	Cuentas 17	masculino	español					Juan Manuel de los, Reyes	
Maria de los Reyes	PERSON	Cuentas 17	femenino						Maria de los, Reyes	Vecino de Guadalajara

Figura 3. Tabla en OpenRefine que muestra los mismos registros que la anterior, pero con las columnas separadas por etiqueta y comentario. Se puede ver cómo en la columna donde únicamente deberían registrarse castas, aparecen oficios (bachiller, Sacerdotes). Fuente: elaboración propia.

También definí qué tipo de información debía ir en los comentarios (campo libre que permite Recogito), como por ejemplo los nombres completos (normalizados con ortografía moderna), divididos por coma entre nombre y apellido para poder separarlos en dos columnas en OpenRefine; o fechas estimadas de nacimiento inferidas a partir de edades declaradas en los documentos. Esta información no forma parte de las etiquetas sino de los comentarios, y puede incluirse tanto en los campos de atributos de las entidades como en los de notas sobre la interpretación de los datos.

El proceso estandarizado de etiquetado lo resumí en el esquema de Recogito para cada uno de los tipos de entidad (*person*, *place*, *event*). Para generar este esquema consideré ejemplos de uso y decisiones sobre cuándo aplicar etiquetas específicas, cómo nombrar entidades nuevas y cómo documentar la información incierta. Este esquema funcionó como protocolo interno para garantizar coherencia durante la anotación.

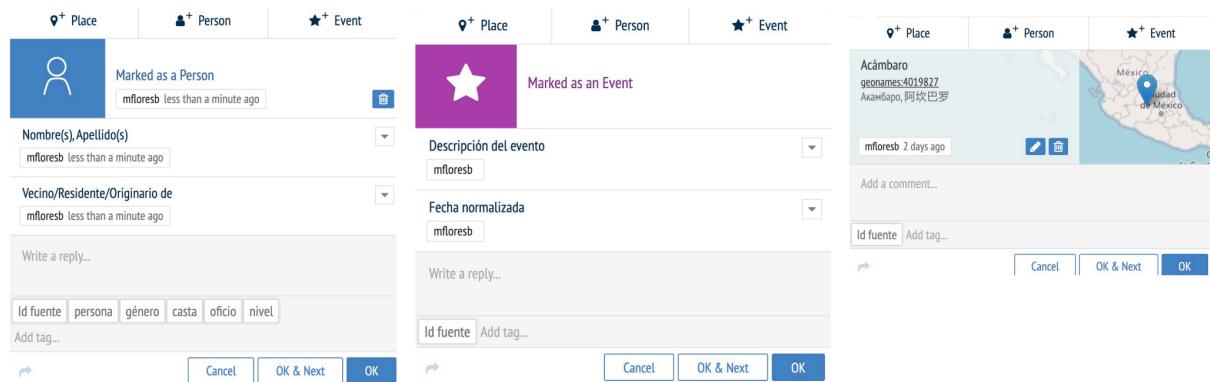


Figura 4. Esquemas de anotación. Fuente: elaboración propia.

Una vez que etiqueté las entidades correspondientes, usé la función ofrecida por Recogito para anotar relaciones. Las etiquetas que se agregan en esta etapa no son controladas, es decir, Recogito no ofrece la opción de agregar un vocabulario para definir relaciones, como sí lo ofrece para las entidades. En este caso, el vocabulario controlado o las reglas de registro de relaciones pueden tenerse en un documento aparte con las definiciones para cada una de las relaciones y siempre consultarse para evitar errores, por ejemplo, saber que se registrará “padre o madre de”, pero no “hijo o hija de”.

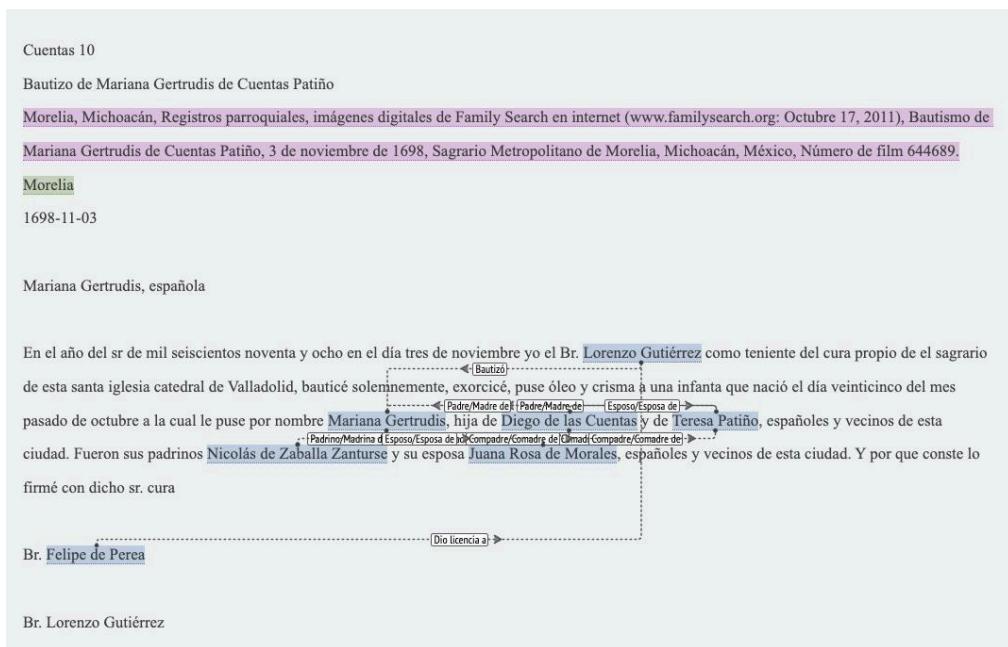


Figura 5. Ejemplo de documento anotado por entidad y con relaciones. El subrayado rosa indica que se trata de un evento, el verde un lugar y el azul una persona.

Fuente: elaboración propia.

Como mencioné anteriormente, al evitar el uso del reconocimiento automático de entidades nombradas (NER), pude controlar manualmente la cantidad de menciones de cada persona, por lo que establecer las conexiones también fue un proceso más limpio y organizado.

Una vez etiquetadas las entidades y las relaciones, Recogito permite descargar los archivos en formato CSV, uno para las entidades y otro para las relaciones. También ofrece descargarlo en un formato compatible con Gephi⁷, un software abierto de análisis de redes sociales. Al seleccionar esta opción, Recogito separa el archivo de las relaciones en un CSV con la lista de entidades y otro CSV con la lista de relaciones entre las entidades. Esta funcionalidad es muy útil si lo que se busca es cargar los datos extraídos de Recogito directamente en Gephi, pero yo necesitaba primero limpiarlos y cargarlos en mi base de datos.

⁷ Disponible en <https://gephi.org/>.

QUOTE_TRANS	TAGS 1	TAGS 2	TAGS 3	TAGS 4	TAGS 5	TAGS 6	COMMENTS 1 1	COMMENTS 1 2	COMMENTS 2
Maria Ferrel	Cuentas 3	nueva persona	femenino				Maria	Ferrer	
Andrés Comparán	Cuentas 3	nueva persona	masculino	casta	Frailes		Andrés	Comparán	
Diego de las Cuentas	Cuentas 4	https://people-things-traces.org/actor/00163	masculino	español			Diego de	Cuentas	Vecino de Valladolid
Theresa Patiño	Cuentas 4	nueva persona	femenino	española			Teresa	Patiño	Vecina de Valladolid
Luis de Salas	Cuentas 4	nueva persona	masculino	español			Luis de	Salas	vecino de Valladolid
Juan Muñoz Xijon	Cuentas 4	nueva persona	masculino	español			Juan	Muñoz Gijón	vecino de Valladolid
Po. Hortis	Cuentas 4	nueva persona	masculino				Pedro	Ortiz	Vecino de Valladolid
Thomas Perez Deza	Cuentas 4	nueva persona	masculino	casta	Sacerdotes	bachiller	Tomás	Pérez Deza	
José de la Sierra	Cuentas 5	nueva persona	masculino	casta	Sacerdotes	doctor	José de la	Sierra	
Diego	Cuentas 5	nueva persona	masculino	español			Diego de	Cuentas Patiño	

Figura 6. Tabla en OpenRefine que muestra el resultado de aplicar las reglas de anotación definidas en el esquema. Fuente: elaboración propia.

Como puede verse en la figura anterior, la mayoría de las personas etiquetadas en Recogito no estaban registradas anteriormente en la base de datos, por lo que aparecen sin una URI. Por el contrario, Diego de Cuentas ya estaba identificado y poseía un número único de identificación. Las siguientes columnas, una vez separadas en OpenRefine, muestran las categorías de etiquetado tal como se definieron y organizaron en los esquemas de la Figura 4: género, casta, ocupación, nivel, nombre(s), apellido(s) y “vecino de”.

Una vez etiquetado el texto que contenía los 23 documentos parroquiales y limpiado el archivo CSV usando OpenRefine, actualicé las tablas correspondientes en la base de datos.

3. CONCLUSIONES

Recogito fue clave para comenzar a estructurar información dispersa. Su interfaz intuitiva y su capacidad de exportación en CSV me permitieron trasladar datos a OpenRefine para limpieza posterior, y de allí a una hoja de cálculo para su eventual integración a la base de datos.

Este ejercicio me enseñó que una herramienta diseñada con un propósito (anotación geográfica y colaborativa) puede adaptarse a otros fines, como la generación de datos históricos estructurados. A la vez, mostró que las decisiones previas, como tener un modelo de base de datos claro, son cruciales para que la anotación sea útil y coherente.

Transformar fuentes narrativas en datos estructurados no es solo un proceso de extracción, sino también de traducción conceptual. Al elegir qué entidades anotar y cómo codificarlas, el investigador redefine los límites de lo que constituye un dato y, en consecuencia, de lo que podrá ser analizado. Esta operación es productiva, pero también selectiva: privilegia relaciones explícitas y medibles, mientras que las sutilezas narrativas como ironías, ambigüedades, o el peso simbólico de un acontecimiento, quedan fuera del registro tabular. Ser consciente de esta pérdida es esencial para no confundir el conjunto de datos con la totalidad de la realidad histórica.

Aunque Recogito no genera redes por sí misma, la información que extraje con su ayuda fue la base para analizar relaciones entre actores en herramientas como Gephi. En este caso, sin embargo, Recogito fue la puerta de entrada: un entorno flexible que me permitió comenzar a pensar en mis fuentes como entidades conectadas y no solo como narrativas lineales. Sin embargo, es importante considerar que la propia estructura de Recogito modela la forma de mirar las fuentes al segmentar el texto en entidades y relaciones discretas, ya que promueve una visión analítica que privilegia los vínculos explícitos y cuantificables. Esto facilita la integración a modelos de datos relacionales, pero también introduce un sesgo hacia la atomización.

Otro de los principales desafíos al usar herramientas de anotación semántica en investigaciones históricas es la necesidad de crear vocabularios que reflejen adecuadamente los contextos del pasado. En este ejercicio, opté por incluir una taxonomía ocupacional propia, basada tanto en los términos originales presentes en los documentos como en vocabularios controlados que se adaptaban al contexto histórico. Esta fue la taxonomía cargada en Recogito que me permitió seleccionar el término adecuado al momento de anotar. Incluye oficios manuales y artísticos (carpinteros, plateros, pintores), cargos eclesiásticos (arzobispos, frailes,

lectores), puestos administrativos (corregidores, alcaldes del crimen), profesiones liberales (abogados, médicos), grados académicos (bachilleres), así como figuras del sistema nobiliario (condes, marqueses, virreyes). También decidí incluir categorías que habitualmente quedan fuera de los registros tradicionales, como personas esclavizadas, criadas o cacahuateros, para visibilizar su papel en los circuitos de trabajo y relación.

La mayoría de los términos para definir los oficios están incluidos en el *tesauro del Getty*⁸; sin embargo, tuve que usar también Wikidata para otros oficios menos relacionados con el mundo del arte, como posiciones administrativas o militares. Por ejemplo, en tres casos distintos, el oficio descrito en el documento no tenía equivalencia con los términos del Getty o de Wikipedia, por lo que tuve que generar mi propia definición y URI interna para ellos. Esta estandarización me permitió mantener consistencia en el etiquetado y facilitar, posteriormente, la limpieza de datos en OpenRefine y su incorporación en la base de datos.

El ejercicio exploratorio realizado con Recogito demuestra que la anotación semántica, cuando se aplica a un corpus cuidadosamente seleccionado y acompañado de un protocolo de etiquetado coherente, es una herramienta potente para transformar documentos históricos en datos estructurados reutilizables. En este caso, el conjunto elegido permitió poner a prueba un flujo de trabajo optimizado que incluyó vocabularios personalizados, estrategias de desambiguación y una jerarquía de etiquetas pensada para garantizar la consistencia de los datos desde la exportación.

Sin embargo, este mismo conjunto también revela los límites de la fuente. Al estar compuesto únicamente por registros parroquiales, la red resultante deja fuera conexiones fundamentales para comprender la historia familiar de Diego de Cuentas. Tal como se mencionó en la

⁸ Disponible en: <https://www.getty.edu/research/tools/vocabularies/aat/>.

introducción, su filiación fue ocultada durante gran parte de su vida por ser hijo natural de un sacerdote. No es sino hasta la lectura de su testamento y de un pleito ante el Tribunal de Bienes de Difuntos que emerge el relato completo: que fue dejado a la puerta de la casa de su tío como niño expósito e incluso reconocido por su padre biológico muchos años después. Nada de esto figura en las actas parroquiales que conforman el corpus, por lo que, si se graficara la red únicamente con esta evidencia, no aparecerían vínculos con Juan Martínez de Alencaster ni con María de Sandoval, que explicarían por qué se le había relacionado con el apellido Alencaster en la historiografía.

Este caso subraya una lección clave: las herramientas digitales como Recogito amplifican nuestra capacidad para sistematizar y visualizar relaciones, pero no reemplazan la interpretación histórica. La elección de qué anotar, cómo etiquetar y qué normalizar determina no solo la forma final del conjunto de datos, sino también las preguntas que podremos responder con él. Es necesario entender, además, que la transformación de fuentes narrativas en datos estructurados no es un proceso neutral, sino una operación interpretativa en la que el investigador deja una huella decisiva. Cada decisión, desde el diseño del modelo de datos hasta la elección de etiquetas o la normalización de un término, implica un equilibrio entre fidelidad a la fuente y coherencia estructural. En este sentido, Recogito funcionó como un espacio intermedio entre el archivo y la base de datos, un laboratorio de experimentación donde el texto aún conservaba su contexto narrativo, pero comenzaba a adquirir la forma tabular que exige el análisis relacional.

Por lo tanto, la estructuración de datos, por rigurosa que sea, debe ir acompañada de una lectura crítica de las fuentes y de un análisis que atienda tanto a lo que dicen como a lo que callan. En última instancia, la fortaleza de este enfoque radica en combinar la precisión de la anotación con la perspectiva historiográfica, para así reconstruir no solo las redes

documentadas, sino también las ausencias que moldean nuestra comprensión del pasado.

REFERENCIAS

- Cabrera, M. (1756). *Maravilla americana y conjunto de raras maravillas, observadas con la direccion de las reglas de el arte de la pintura en la prodigiosa imagen de Nuestra Sra. de Guadalupe de México*. Imprenta del Real y más antiguo Colegio de San Ildefonso. <http://archive.org/details/maravillaamerica00cabr>
- Diego-Fernández Sotelo, R. y Mantilla Trolle, M. (Eds.). (2003). *La Nueva Galicia en el ocaso del imperio español: los papeles de derecho de la audiencia de la Nueva Galicia del licenciado Juan José Ruíz Moscoso su agente fiscal y regidor del Ayuntamiento de Guadalajara, 1780-1810*. El Colegio de Michoacán.
- Flores Barba, M. L. (2009). El obrador de la familia Cuentas en Guadalajara. *Anales del Instituto de Investigaciones Estéticas*, 31(95), 69–84.
- Flores Barba, M. L. (2014). *Diego de Cuentas, pintor de entresiglos en la Nueva Galicia (1654-1744)* [Tesis de maestría, Universidad Nacional Autónoma de México]. http://ru.atheneadigital.filos.unam.mx/jspui/handle/FFYL_UNAM/5342
- Flores Barba, M. L. (2025). *From Historiography to networks: A digital approach to colonial Mexican painters (1680-1730)* [Tesis de doctorado, Western University]. <https://ir.lib.uwo.ca/etd/10894>
- Gila Medina, L. (2015). Aproximación a la vida y obra del pintor novohispano Antonio de Torres (1667-1731) y estudio de una serie inédita mariana del convento de la Encarnación de Granada de franciscanas clarisas. *Anales del Museo de América*, 18, 82–113.
- González Franco, G., Olvera Calvo, M. del C. y Reyes y Cabañas, A. E. (1994). *Artistas y artesanos a través de fuentes documentales*. (Vol. I). INAH.

- Gutiérrez Haces, J. (Ed.). (1997). *Cristóbal de Villalpando: catálogo razonado, 1549-1714*. Fomento Cultural Banamex, CONACULTA, Instituto de Investigaciones Estéticas.
- Mayorga Chamorro, J. I. (2019). *Fact. Mexico. El pintor novohispano Pedro López Calderón (activo ca. 1681- ca. 1734)* [Tesis de doctorado, Universidad de Málaga]. <https://riuma.uma.es/xmlui/handle/10630/19051>
- Mota Padilla, M. A. de la. (1870). *Historia de la conquista de la provincia de la Nueva Galicia*. Impr. de gobierno, á cargo de J.M. Sandoval.
- Orendáin, L. (1951). El pintor Diego A. de Cuentas. *Anales del Instituto de Investigaciones Estéticas*, 5(19), 75-85. <https://doi.org/10.22201/iee.18703062e.1951.19.520>
- Pasin, M., y Bradley, J. (2015). Factoid-based prosopography and computer ontologies: Towards an integrated approach. *Digital Scholarship in the Humanities*, 30(1), 86-97. <https://doi.org/10.1093/llc/fqt037>
- Ramírez, M. V. (2009). *Ignacio Berben: un pintor del reino de la Nueva Galicia, siglo XVIII*. Amate Editorial.
- Ramírez Montes, M. y Luckie, G. (1993). *Catálogos de documentos de arte. Archivo de Notarías de la Ciudad de México. Protocolos I*. Universidad Nacional Autónoma de México, Instituto de Investigaciones Estéticas.
- Rubial García, A. (2010). *El paraíso de los elegidos. Una lectura de la historia cultural de Nueva España*. Fondo de Cultura Económica.
- Toussaint, M. (1965). *Pintura colonial en México*. Imprenta Universitaria, Universidad Nacional Autonoma de Mexico.
- Vargas Lugo, E., Victoria, J. G. y Curiel, G. (1991). *Juan Correa: Su vida y su obra. Cuerpo de documentos*. (Vol. III). UNAM, Instituto de Investigaciones Estéticas.
- Vázquez, G. S. O. (2019). Mediatisación y digitalización del patrimonio cultural. El caso de la Mediateca del Instituto Nacional de Antropología e Historia. *Revista de Humanidades Digitales*, 4, 69-85. <https://doi.org/10.5944/rhd.vol.4.2019.25110>