

DE LA SEÑAL A LA BASE DE DATOS.
APROXIMACIÓN A LA TRANSFORMACIÓN
DE FUENTES HISTÓRICAS EN BASES DE DATOS
DE GRANULARIDAD FINA¹

FROM SIGNAL TO DATABASE. APPROACH TO THE
TRANSFORMATION OF HISTORICAL SOURCES INTO
FINE-GRAINED DATABASES

Pablo PRESUMIDO CASADO* & Jean Pierre DEDIEU
Universidad del País Vasco & CNRS Institut d'Asie Orientale
presumido@ehu.eus ORCID: <https://orcid.org/0000-0002-8251-6507>
jeanpierre.parpailou@gmail.com ORCID: <https://orcid.org/0000-0002-5684-4718>

RESUMEN: Tomando Fichoz como ejemplo para la elaboración de una base de datos enfocada a la investigación histórica, mostraremos algunos elementos clave para la creación de esta clase de bases de datos. Un primer apartado teórico expondrá los dos niveles requeridos. La fase inicial, implementada durante la etapa de introducción de datos apoyada en una ontología resumida que proporcione información segmentada en una serie de acciones descritas de forma sistemática basándose en cuatro criterios puramente formales: quién, qué, dónde, cuándo y, en algunos casos, con quién. Una segunda etapa, introducida posteriormente, anotando

Recibido: 24-10-2024; Aceptado: 26-03-2025; Versión definitiva: 11-05-2025

* Proyecto de Investigación del Ministerio de Ciencia e Innovación del Gobierno de España PID2020-114496RB-I00, titulado: «Disrupciones y continuidades en el proceso de la modernidad, siglos XVI-XIX. Un análisis multidisciplinar (Historia, Arte, Literatura)». Grupo de investigación del Gobierno Vasco «Sociedades, Procesos, Culturas (siglos VIII a XVIII)», ref. IT1465-22.

1. El presente artículo ha tomado como base un primer borrador elaborado por Jean Pierre Dedieu para la revista francesa *Histoire & Mesure* en calidad de introducción a un monográfico sobre bases de datos para la investigación histórica en el que también han participado los investigadores Silvia Marzagalli, Albane Cogne y Gunnar Knutsen. Dicho monográfico se encuentra actualmente en preparación para su publicación y cuenta con los permisos necesarios para la presente publicación. Recomendamos al lector de este artículo su lectura una vez publicado. El presente estudio, sin embargo, ha sido traducido, reestructurado y ampliado, tanto en su parte teórica como práctica, dando, finalmente, un producto diferente al original. Así, ha pasado de ser una breve introducción de diversos trabajos a convertirse en un planteamiento sólido sobre cómo deben estructurarse las bases de datos destinadas al estudio de la historia que tiene valor por sí mismo. Los ejemplos que se exponen en la segunda parte del trabajo son totalmente novedosos con respecto al texto original, así como la práctica totalidad de la parte teórica.

Copyright: © Editorial Universidad de Sevilla. Este es un artículo de acceso abierto distribuido bajo los términos de la licencia de uso y distribución Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 (CC BY-NC-SA 4.0)

y desarrollando si fuera necesario cada una de las acciones descritas en la primera capa, utilizando en su caso ontologías más complejas que permitan descripciones precisas. La segunda mitad del trabajo muestra ejemplos de la segmentación de la información archivística y su mejora exponencial en el tratamiento de los datos, su manejo y posterior análisis.

PALABRAS CLAVE: base de datos; acciones; datos; sistematización; Fichoz.

ABSTRACT: Taking Fichoz as an example for the development of a database for historical research, some keys for the creation of this type of database will be presented. A first theoretical section will expose the two levels for the development of a base of these characteristics. The initial phase, implemented during the data entry stage supported by a summarized ontology that provides information segmented into a series of systematically described actions based on four purely formal criteria: who, what, where, when and, in some cases, with who. A second stage, introduced later, annotating and developing, if necessary, each of the actions described in the first layer, using, where appropriate, more complex ontologies that allow precise descriptions. The second half of the work shows examples of the segmentation of archival information and its exponential improvement in data processing, management and subsequent analysis.

KEYWORDS: database; actions; data; systematization; Fichoz.

1. INTRODUCCIÓN

En las siguientes páginas se exponen los criterios que los autores de este artículo consideran necesarios a la hora de elaborar una base de datos para la investigación histórica. Varios autores, entre los cuales destacan en última instancia Díaz Ordóñez, Rodríguez Baena y Yun-Casalilla, expusieron las pautas técnico-informáticas que los investigadores deben tener claras a la hora del desarrollo de una de estas bases, tales como las ventajas del lenguaje NoSQL sobre el SQL². Nosotros no nos fijamos en la selección de un paquete informático de tal o cual tipo, sino en el proceso de análisis de la información proporcionada por las fuentes para reducirla a un modelo compatible con el proceso informático, fuera cual fuera el instrumento usado. Nuestra intención es situarnos desde el punto de vista de las necesidades específicas que el investigador puede encontrar a la hora de analizar la documentación, haciendo algunas proposiciones independientes al sistema informático que se decida emplear. Tal y como afirmaba Ana Paulina Malavassi, es necesario que estas bases ofrezcan la posibilidad de unificar la información proveniente de diferentes fuentes, la agilización en la consulta de datos, la posibilidad de introducir cambios en la estructura de los datos y su actualización y la depuración de los mismos, eliminando la posibilidad de duplicados³. ¿Cómo aplicar estas

2. Díaz-Ordóñez *et al.* 2023.

3. Malavassi Aguilar 2012, p. 195.

pautas generales al desarrollo de una nueva base o emplearlos para seleccionar una u otra base ya existente para desarrollar una investigación histórica? Esa es la cuestión que pretendemos aclarar.

Teniendo presentes dichas reglas fundamentales, expondremos los criterios específicos que el investigador ha de considerar en el desarrollo de la misma según el criterio razonado de los autores de este trabajo. La intención final es que el lector concluya por sí mismo sobre la posible utilidad de una base de datos para su trabajo teniendo en cuenta el coste en dinero y en tiempo generado por la implementación en ella de los criterios esenciales mostrados para un diseño que le permita eficacia durante el procesamiento de los datos.

La primera parte se centra en la exposición teórica de las propuestas más relevantes, siempre enfocadas a las necesidades específicas de la investigación histórica y no tanto a una reflexión general sobre bases de datos. La segunda, en ejemplos basados en la base de datos semirrelacional⁴ Fichoz. La selección de esta base no es aleatoria. Por un lado, a diferencia de la mayoría, es accesible a cualquier investigador que lea estas páginas a través de internet (<https://actoz.db.huma-num.fr/fmi/webd>). Por otro, es una de las bases de datos históricas más antiguas, con una vida ininterrumpida de más de 30 años que ha servido para decenas de proyectos sobre las temáticas más variadas, usando una gran diversidad de fuentes. No obstante, no se trata de promocionar Fichoz, tampoco de comparar esta base con otras existentes, sino de exponer prácticamente sus fundamentos, de ya demostrada eficacia, de tal forma que sean tenidos en cuenta para el desarrollo de una base de datos por quien desee hacerlo, con clara conciencia de las posibilidades y limitaciones que ofrece un sistema de este tipo y los criterios elaborados para su desarrollo.

2. ¿QUÉ ESPERAR DE UNA BASE DE DATOS?

El reciente desarrollo de las metodologías en ciencias sociales impone a la investigación histórica un seguimiento individualizado, pormenorizado y multidimensional de los actores y su comportamiento en el tiempo⁵. Ello requiere centrar la investigación no sólo en la descripción de equilibrios globales entre clases de actores, sino también en la agregación de comportamientos individuales. Estos pueden ser descritos como procesos adaptativos, o sea una sucesión de modelos de conducta progresivamente refinados para adaptarse al contexto relacional corriente, que van siendo construidos por cada uno de los actores a base de pruebas y errores en un proceso evolutivo que se inscribe en el tiempo y cuyas distintas

4. Las bases de datos relacionales se basan en el empleo de diferentes tablas divididas en filas y columnas, todas ellas relacionadas entre sí. Sánchez 2004, p. 12. Fichoz se apoya en el paquete FileMaker, que no era en su origen relacional, sino una base de datos plana, pero que se agregó con el tiempo rasgos relacionales, lo que le permite combinar la sencillez y la capacidad de evolución de las bases planas, con la potencia de las relacionales.

5. Dedieu 2017.

etapas es importante recoger para entender la configuración final: el observador no puede limitarse únicamente a considerar los éxitos y descartar los errores. Dicho proceso alberga por necesidad una complejidad notable que se deriva del gran número de configuraciones posibles que resultan de las múltiples opciones que el actor tiene en cada momento para elegir diferentes alternativas de acción; que se deriva también del hecho de que estas elecciones dependen a su vez de numerosos factores (convenciones sociales implícitas, elementos de la historia personal de los actores, por ejemplo) que no se explicitan en la situación inicial y que el trabajo del investigador tendrá que ir desentrañando uno por uno⁶.

La multiplicidad de alternativas que se ofrecen a cada actor en cada uno de los momentos históricos –determinadas por factores internos y externos– dificultan la posibilidad de predicción del investigador. En buena medida, también obstaculizan la comprensión del fenómeno analizado. La confusión crece aún más cuando el investigador debe afrontar las relaciones sociales desarrolladas en épocas histórica pasadas –incrementándose dicha dificultad a medida que se retrotrae en el tiempo– debido a que no es conocedor de primera mano de las herramientas colectivas que utilizaron los actores que estudia para gestionar las relaciones sociales. Existe, por tanto, una brecha evidente entre las claves sociales y culturales interiorizadas por el investigador y aquellas mismas claves comprendidas por los investigados. Solo un acercamiento de la perspectiva del investigador al mundo del investigado le capacita para una correcta interpretación de los datos.

Sin duda, son aquellas claves las que vertebran, estructuran y acaban produciendo la información en crudo que arrojan las fuentes de archivo. Estos documentos, que pasaremos a llamar la “señal”⁷, son el punto de partida de todo historiador.

6. La necesidad de priorizar el estudio del individuo –el actor– al hacer historia social se explica en Dedieu 2021. La historia social tradicional se centraba en los actores colectivos, quitando toda la importancia al individuo. Sin embargo, hoy se reconoce la importancia del individuo como primer eslabón de la acción colectiva. La actuación individual no viene únicamente regida por la pertenencia a una u otra “clase” predeterminedada, sino que en ella juegan un papel destacado factores como la racionalidad y la utilidad que dicho individuo reconoce en su actuación, a la par que el estatus que disfruta dentro de cada uno de los diversos “mundos” sociales en los que se inserta (la familia, el oficio, la pertenencia a una congregación religiosa, el testamento, etc.). Todo ello, además, solo puede ser entendido dentro de un contexto. Por tanto, es necesario un análisis del individuo y las relaciones desarrolladas por aquel para comprender realmente las causas de unas y otras decisiones en cada momento determinado, lo cual solo es posible a través del análisis de grandes volúmenes de documentación.

7. No será demás aclarar unas distinciones conceptuales que sacamos libremente de Luciano Fioridi. Las fuentes entregan una señal, que no es el fenómeno estudiado, sino una huella del mismo, huella parcial, derivada, mezclada con elementos irrelevantes que constituyen lo que la teoría de la información llama «ruido». El dato es esta misma señal puesta en forma, limpiada del ruido, reconstituido en cuanto se pueda los elementos destruidos por el instrumento que la transmitió (en historia, este instrumento se llama la fuente). Los datos organizados en series conforman una información. La información, inserta en una estructura cognitiva externa a sí misma constituye un conocimiento. Fioridi 2013, 2019. Así una inscripción latina. El objeto material en que figura es la señal. Una vez las abreviaturas desarrolladas, las partes que faltan restituidas conforme a las reglas del arte sobre las que hay consenso de los especialistas, se hace dato. Las inscripciones referentes al mismo individuo, una vez montadas en una serie, por ejemplo, cronológica, nos dicen que el interesado fue primero tribuno en España, luego legado en Siria y por fin senador en Roma: es una información. Determinar que esta información constituye un *cursus honorum* exige la inyección del concepto de *cursus honorum*, un

Sin embargo, el conocimiento de esta señal emitida documentalmente por las sociedades antiguas formaba parte de su propio sistema interpretativo, por lo que no puede ser nítidamente percibida por el historiador sin un análisis previo. Más bien, la comprensión de dichas claves es el fruto deseado de la investigación una vez concluida. Consecuentemente, existe un riesgo altísimo de causar malentendidos en la comprensión de las fuentes si no se toman las medidas oportunas. Interpretaciones apresuradas de la “señal” limitarían inevitablemente su hermenéutica, dándole una nueva semántica a la misma que sería generalmente creación del propio investigador y no comprensión razonada de lo analizado. Un ejemplo de tales distorsiones, claramente vigente en la historiografía española actual, es la interpretación que se hace sobre la venalidad. Si bien en el contexto de finales del siglo XVII podría comprenderse como el intercambio de servicios a cambio de gracias, aunque dichos servicios fuesen en dinero, actualmente se malinterpreta como corrupción sistemática⁸. Claro está que esta clase de interpretaciones obedecen a nuestra propia manera de entender la sociedad y no a una correcta interpretación de los datos emitidos por la señal.

Desde el punto de vista más práctico, las nuevas metodologías ya referidas arriba requieren inexcusablemente del empleo de bases de datos informatizadas. Solamente a través de ellas resulta posible sistematizar la ingente masa y la complejidad de los datos para reducirlos a acciones como estrategia que permita su posterior análisis⁹. Las condiciones individuales del investigador, sumadas a la cantidad de información que debe recopilar incluso para el estudio de un solo caso, hace poco realista la recopilación manual –en adelante “proceso analógico”– de

elemento que no figuraba en la información. Tal inyección transforma la información en conocimiento, lo que permite reconstruir por inferencia posibles datos desaparecidos e integrar la inscripción en una teoría general del imperio romano. La distinción, en la metodología histórica, entre la señal (que llama “huella”), y el dato, es el tema fundamental de Seignobos 1901, a quien poco se puede añadir. Una base de datos tiene, en rigor, que atenerse los datos, y permitir al mismo tiempo la elaboración de información a partir de aquellos; información creada por el investigador fuera de la base, en función de las necesidades de su investigación, combinando los datos. En cuanto a los elementos que sirven a la producción de conocimiento, dependen altamente de la iniciativa del investigador, de sus presupuestos y de una serie de factores personales, lo que llevaba Seignobos a dudar de que la historia fuera una ciencia. No le seguiremos en esta vía, pero sí sacaremos de sus observaciones el que una base de datos no tiene que integrar elementos de conocimiento en el sentido que damos aquí a esta palabra, so pena de impedir cualquier uso colectivo de sí misma.

8. Artola Renedo; Dedieu 2011, p. 4. Dedieu 2011.

9. En el año 2013, se publicó un artículo sobre la base de datos Fichoz en el que ya se señalaba que la gestión de la información extraída de la señal era mucho más eficaz a través de dicha base de datos. En un primer momento, se trataba únicamente de la recopilación de datos sobre carreras individuales en la administración de la Monarquía española del siglo XVIII. La enorme cantidad de información necesaria había impedido anteriormente llevar a término dicho estudio siguiendo únicamente estrategias analógicas. Esto había desanimado al profesor Didier Ozanam y a su equipo de investigadores, entre los que constaba Janine Fayard, a continuar la explotación sistemática de los fondos de Simancas, pues la información se almacenaba en cartulinas correspondientes a cada nombramiento. Cuando se llevó a cabo el proceso de informatización –de ahí el nombre de Fichoz, Ficheros-Ozanam– se vio que existían grandes vacíos documentales, a pesar del abundante material físico recopilado, inmanejable en su forma analógica. Ya en formato digital, no hubo ningún problema en reanudar las campañas y completar el trabajo, mucho más accesible a los investigadores. Dedieu 2013, p. 194.

todos los datos necesarios, y menos la de las anotaciones necesarias para situar cada concepto, cada palabra en el contexto de su tiempo. El procedimiento analógico impide, además, una rápida reconfiguración o corrección de los datos cuando esta es necesaria¹⁰.

Por tanto, sólo la informática y el desarrollo de herramientas bien diseñadas permite organizar el trabajo de forma eficaz según los nuevos parámetros epistemológicos. Esta metodología exigirá del investigador la convergencia en un espacio homogéneo, alrededor de un número *a priori* indeterminado de actores, del conjunto de datos que describen sus comportamientos. Sin embargo, gracias a la flexibilidad que aporta el diseño e implementación de una base de datos bien planificada desde un inicio para el proceso de recogida de datos, la incorporación de datos inéditos hasta el momento, descubiertos a medida que avanza la investigación, no supondrá ningún problema. La aparición de nuevas informaciones inesperadas, con un orden confuso y en fuentes ajenas a lo esperado es algo habitual, incluso natural¹¹. Por ello, es preciso disponer de herramientas que nos permitan incorporarlas eficazmente a nuestros estudios. La informática nos da la posibilidad de desarrollar las herramientas necesarias –las bases de datos– para agregarlas al sistema de la misma forma en que se incorporó el primer dato registrado, independientemente de la cronología o fuente de la que se rescate. Cada dato aportará un nuevo plano de la realidad del actor, una nueva posibilidad, en fin, una nueva acción¹².

10. En el aspecto más práctico, por ejemplo, la recopilación analógica –i. e. no digital– de los datos no permite insertar informaciones nuevas a medida que van siendo descubiertas. Pongamos como ejemplo fichas manuales en las que se recopila toda la información biográfica de un personaje. Una vez recogido un dato para un año concreto, resulta difícil y tedioso introducir nueva información sobre una cronología anterior por la falta de espacio físico en dichas fichas. Aunque resulte una obviedad, cabe resaltar la ineeficiencia de un sistema como el descrito para los estudios actuales. Concretamente, o bien se renuncia a la introducción del nuevo dato referido o se procede a elaborar una ficha completamente nueva con la subsiguiente pérdida de tiempo.

11. Los investigadores Manuel Díaz-Ordóñez, Domingo Savio Rodríguez Baena y Bartolomé Yun Casalilla proponen en un reciente estudio de 2023 el empleo de bases de datos que emplean el lenguaje informático NoSQL. Este lenguaje informático es diferente al lenguaje SQL empleado por Fichoz y las bases de datos relacionales en general. El lenguaje NoSQL permite una mayor flexibilidad en el tratamiento de los datos, tal y como exponen para la base *AtlantoCracies* que están elaborando para el estudio de las noblezas atlánticas de época moderna. Entre otras cuestiones a señalar, permite la modificación de campos en la estructura de dichas bases de una forma más sencilla que en las bases con lenguaje SQL. Sin embargo, nosotros no queremos centrarnos en esta cuestión, precisamente, porque ya se trata en el artículo mencionado de una forma detallada. Más bien, pretendemos dar una idea lo más práctica posible acerca de los criterios específicos a tener en cuenta en el diseño de una base de datos para quien esté valorando su uso, de tal forma que tenga una noción lo más realista posible acerca de lo que puede esperar de la misma. De tal forma, ambos artículos vendrían a complementarse, puesto que mientras uno ofrece una visión técnica que debe ser considerada por el investigador, esta aportación se refiere a los criterios a tener en cuenta una vez se ha decidido en qué lenguaje informático se pretende desarrollar dicha base de datos. Díaz-Ordóñez *et al.* 2023. También se recomienda la revisión de la intervención de Adrián Díaz “Migración desde sistemas relacionales a sistemas NoSQL para una base de datos históricas: AtlantoCracies” que saldrá publicada en las actas del Encuentro Internacional de Jóvenes Investigadores en Historia Moderna, celebrado en Jaén-Úbeda entre los días 24-26 de abril de 2024. En este mismo sentido, el artículo Ortín-Ibáñez *et al.* 2023.

12. El diseño de una base de datos que siga el criterio de la “acción” como eje principal permitirá al investigador tratar cualquier acontecimiento en acciones uninominales (nacimiento, muerte, etc.)

El principio de homogeneidad en la recogida de los datos, que exige recolectarlos todos y prohíbe seleccionarlos eliminando algunos, supuestamente irrelevantes, en la primera etapa del trabajo ya, sólo se puede implementar en la práctica por la informática y es indispensable para dar cuenta cabal de lo observado.

La abundante masa de información percibida de las múltiples fuentes consultadas por el investigador envía señales diferenciadas entre sí, formateadas de distinta manera, dependiendo de muchos factores, tanto del lado del productor o del receptor. Tales señales deben, a nuestro entender, transformarse y sistematizarse mediante la base de datos para que adquieran un carácter uniforme, una única forma, y someterse a una homogeneización posible gracias al tratamiento por el ordenador eliminando entre otras operaciones todos los elementos duplicados que inevitablemente nos presentarán las fuentes¹³. Paralelamente, el uso de la tecnología digital debe permitirnos, de manera rápida y sencilla, la construcción de nuevos datos, la corrección de aquellos introducidos anteriormente y la posibilidad de dejar de lado, provisionalmente, los que ya no deseemos insertar en el análisis en curso, de una manera muy flexible; todo ello, sin embargo, sin requerir de una restructuración de la base de datos desde su raíz, sino mediante un proceso sencillo y de escasos minutos. Como puede apreciarse, herramientas de tales características son todo lo contrario de lo que ofrece la práctica clásica de las bases de datos, apoyada en el desarrollo previo de corpus documentales, ontologías¹⁴ y esquemas rígidos de organización de la información¹⁵. La necesaria flexibilidad que pedimos es el desafío mayor que la investigación histórica plantea al informático y este es, en nuestra opinión, el principal obstáculo que se opone a la generalización del uso sistemático de las bases de datos en la investigación histórica.

o binominales (actor 1 se relaciona con actor 2 para tal o cual cuestión). Dedieu 2004, p. 6. Como ejemplo de eficacia de este criterio, debemos señalar la adecuación del sistema Fichoz para los investigadores que la han empleado. Daremos algunos ejemplos prácticos de la aplicabilidad de este criterio en la segunda parte del trabajo.

13. La base de datos nos permite comprobar en tan solo unos segundos los datos que ya están introducidos. La misma comprobación analógica puede demorarse mucho más, dependiendo de la calidad de los registros manuales creados por el investigador. Sin embargo, cuando aquél no tenga a mano dichos registros –por ejemplo, durante una visita a un archivo– no podrá comprobar los datos que ya tiene. Por ello, deberá volver a anotarlos perdiendo buena parte del tiempo de archivo en ello y, posiblemente, contribuirán estos datos a generar confusión durante el análisis ya que algunos aparecerán duplicados, triplicados o cuadruplicados. Esto puede llevar a interpretaciones erróneas del fenómeno analizado.

14. “En informática y en ciencias de la información, una ontología es el conjunto estructurado de los términos y conceptos que representan el sentido de un campo de informaciones, sea por medio de metadatos, de un conjunto de nombres o de los elementos que articulan un campo de conocimientos. Una ontología constituye, de por sí, un modelo de datos representativo de un conjunto de conceptos en un campo concreto, así como de las relaciones entre estos conceptos. Se usa para razonar acerca de los objetos del campo interesado”. «Ontologie (informatique)» 2022 (traducción por los autores de la entrada). La llamada «inteligencia artificial» no resta valor alguno al concepto de ontología. La inteligencia artificial se limita a proyectar sobre los datos las ontologías que elaboró el paquete que lleva a cabo la operación en el curso de su aprendizaje, con una eficiencia fascinante, pero al precio de un enorme riesgo, especialmente notable en la investigación histórica, ya que tales ontologías pueden no corresponder a las que usaron los autores de la fuente en la elaboración de la misma.

15. Díaz-Ordóñez *et al.* 2023.

2.1. Cinco propuestas

Existen varias opciones para sortear las contradicciones planteadas en el epígrafe anterior. Analicemos como ejemplo un instrumento concreto, que tiene una gran virtud: funciona desde hace más de treinta años¹⁶ al servicio de problemáticas y fuentes de lo más variadas, sin haber necesitado nunca una reestructuración en profundidad¹⁷. Se trata de una familia de bases de datos conocidas colectivamente por el nombre de Fichoz. Este sistema se fue desarrollando gradualmente en los equipos del CNRS desde 1989, primero en la Maison des Pays Ibériques (Bordeaux), luego en LARHRA (Lyon) y finalmente en el IAO (ENS Lyon)¹⁸. Más de 200 investigadores franceses, chilenos, españoles, portugueses, italianos, alemanes y tunecinos lo han utilizado eficazmente¹⁹. En cualquier caso, no se trata aquí de presentar a Fichoz, sino de exponer los principios en los que se basa, principios extrapolables a cualquier otro sistema de información, los únicos, en nuestra opinión, capaces de garantizar la movilización informática eficaz de documentos históricos. Lo haremos en cinco puntos.

A) La primera propuesta establece que la respuesta al problema con que nos enfrentamos no consiste en una remodelación del sistema informático, sino en un mejor análisis de los datos. El ordenador, en el estado actual de la tecnología, es incapaz de improvisar por sí mismo fuera de esquemas previamente planificados e implementados, lo que le resta flexibilidad y lo invalida ante situaciones

16. Tal y como señalaba Peter Cornwell, la conservación de los datos generados a lo largo de una investigación histórica tendría que ser un motivo de profunda inquietud para la comunidad científica. Estos datos, que tienen un alto interés en sí mismos, son además necesarios para la verificación independiente de los resultados científicos obtenidos y, finalmente, publicados. Muchas de las bases de datos y, en general, herramientas para el análisis y estudio histórico, acaban siendo inaccesibles en el término de pocos años por la obsolescencia de los programas empleados y la escasez de recursos invertidos en su salvaguarda. En gran medida, este problema responde a la duración finita de los proyectos de investigación en los que se adscriben, tras los cuales, no se les dota de la infraestructura digital necesaria para su pervivencia. Cornwell s. f., p. 2. Frente a este considerable problema, Fichoz puede ofrecer al investigador un grado elevado de seguridad en la pervivencia de sus datos, como lo demuestra su pervivencia durante más de 30 años, la sencillez de su estructura de conjunto que permite su transcripción fácil bajo otro sistema (ya sufrió dos transcripciones de este tipo en su vida) y al hecho de estar alojada en el servidor Humanum del CNRS, que garantiza una preservación a muy largo plazo.

17. Uno de los mayores problemas a los que se puede enfrentar una base de datos es la necesidad de reestructuración por el cambio de los intereses de los investigadores o la necesidad de añadir un criterio de investigación antes no valorado. Díaz-Ordóñez *et al.*, 2023 p. 1016. El hecho de que Fichoz no haya sufrido una gran reestructuración desde su creación hace interesante valorar los criterios que emplea para el procesamiento de sus datos.

18. Institut d'Asie Orientale, UMR 5064 (CNRS, ENS Lyon, IEP Lyon, Université Louis Lumière Lyon 2).

19. Varios artículos sobre Fichoz nos muestran el gran desarrollo de esta base gracias al trabajo colectivo de los numerosos investigadores que la han empleado para sus investigaciones. Su sección referente al mundo ibérico (hay otras sobre China y sobre la historia del transporte marítimo) en 2005, constaba de unos 250.000 registros correspondientes a más de 43.000 individuos diferentes (Dedieu 2004, p. 9); en 2011, eran 400.000 registros de 70.000 individuos (Dedieu 2013, p. 10); a día de 14 de abril de 2025, son exactamente 673.369 registros. Estos datos muestran la capacidad de esta base para el trabajo colaborativo, así como la pervivencia de los datos de los primeros investigadores que la usaron en la actualidad.

imprevistas. La inteligencia artificial, al menos hasta que esté más avanzada, no cambia nada al meollo del problema: si el sistema no ha sido “educado” en el universo que se está procesando, no dará resultados, o peor, dará resultados equivocados. Queda por determinar si, aún en el caso que haya sido “educada” en un campo, extraerá de los datos conocimientos realmente nuevos. Va a lo más probable, cuando lo propio de un conocimiento nuevo es precisamente la improbabilidad. Eso sí, la inteligencia artificial suaviza y acelera considerablemente algunas fases de la puesta en forma de los datos que no exigen más que conocimientos mecánicos, en campos ya perfectamente trabajados. Ya que no podemos resolver el problema por vía del factor informático, es el otro factor, el procesamiento de los datos, sobre el que debemos apostar, según las líneas que exponemos a continuación.

Ante todo, es necesario aislar cada dato dentro de un marco formal siempre “idéntico” cuya uniformidad haga posible que la máquina lo manipule independientemente de su contenido y permita recuperar la información que contiene a niveles agregativos más altos sin perder el detalle del dato individual. Solo así se logra generar el máximo número de recombinaciones entre los elementos conocidos para seleccionar las más aclaradoras. Para ello, es indispensable mantener inalterable una estructura formal única diseñada para encapsular el dato. Implementar semejante uniformización formal, y mantener aislado y encapsulado cada dato, respetar sistemáticamente estas características formales, no significa que tales características deban trasladarse al análisis de su contenido. La encapsulación del dato no es sino una forma, una herramienta, una estrategia, para el manejo de grandes cantidades de información, permitiendo una comprensión final total del conjunto de dichos datos. Comprensión cimentada en una descomposición fina de los datos en elementos cada vez más menudos, de granularidad cada vez más fina, que permita al investigador la construcción de su aparato interpretativo (véase un ejemplo en el apartado “Hacia el pensamiento por acciones” en la segunda parte del presente texto). Esta operación de disagregación y encapsulamiento puede denominarse como atomización de los datos.

Por granularidad fina queremos decir que los elementos que componen la base de datos se reducen al tamaño más pequeño posible, compatible con la descripción de los objetos que reportan, estandarizados en su presentación para ser accesibles en el mismo espacio informático, dotados de los metadatos necesarios para su procesamiento. O sea que el tamaño requerido de los granos de información varía en función del uso al que aquellos se destinan. Sin duda, la investigación es el sector más exigente en este campo. El análisis implica una descomposición cada vez más fina de los fenómenos estudiados para una comprensión más profunda de lo investigado. La base debe permitir la división de la información originalmente extraída de la fuente en forma de bloques grandes, en granos más finos si el análisis así lo requiere, sin tener que cambiar la estructura de la base.

B) La segunda proposición establece que los datos atomizados son acciones llevadas a cabo por actores. Un actor se describe por las acciones que realiza o que sufre. Por tanto, la base de datos debe consistir fundamentalmente en la

descripción de un conjunto de acciones formalizadas de la manera siguiente: un actor (quién), realiza o sufre una acción (qué), en un lugar (dónde), en un momento dado (cuándo), posiblemente en relación con otro actor (con quién). Estas deben ser las cinco funciones en que se base la recogida de datos, en inglés las cinco “W”²⁰. Igualmente, tomamos el término “actor” en el sentido de Bruno Latour: cualquier instancia que interviene en un conjunto de relaciones entre individuos o entre grupos. En este sentido, por poner un ejemplo famoso desarrollado por Latour, un barco es un actor, al igual que lo es su capitán²¹. Esta estrategia simplifica radicalmente la base de datos. El trabajo del investigador consiste, por tanto, en segmentar las fuentes en secuencias representativas de acciones, según el modelo de las cinco “W”, hasta agotarlas, según el grado de granularidad que considere necesario para la tarea del momento. Es libre de variar el grado de granularidad, ya sea agrupando elementos finos en elementos más extensos, o añadiendo entradas adicionales que dividan los elementos existentes en elementos más finos. En cualquier caso, la agrupación o separación de elementos no debe suponer el reinicio de la sistematización de la señal proporcionada por la fuente, sino apoyarse en las estructuras de la base ya construidas.

Todo ello hace posible procesar en una misma base de datos, bajo una misma estructura formal cualquier información, sea cual sea su forma inicial o fuente original. Esto permite al investigador construir paulatinamente sus datos a partir de la señal guiándose siempre por las reglas de la hermenéutica propias de su disciplina. De hecho, también en la era pre-informática el investigador trabajaba así cuando elaboraba el corpus documental sobre el que basaría sus conclusiones. Semejante continuidad es fruto de la prioridad que damos al elaborar nuestras bases de datos a la perspectiva histórica, no a la informática (ver esquema 02). El proceso requiere una intervención masiva del investigador, que literalmente amolda la materia que entregan las fuentes a las necesidades de su investigación. ¿Será por ello un proceso abierto a la arbitrariedad? La experiencia lograda de Fichoz muestra que no es así, que investigadores que trabajaron por separado sobre los mismos documentos obtuvieron una notable convergencia. Lo que de paso subraya la solidez de las reglas de la metodología histórica en la que todos se apoyaron.

C) La tercera proposición establece que los datos descritos anteriormente deben extraerse de las señales proporcionadas por las fuentes y ser sistematizados en la base, inicialmente, de la forma más fiel posible al original, respetando la manera en que se nos presentan, conformados en la base por una ontología mínima, sin otros vínculos entre ellos que su sucesión cronológica (dada por el factor “cuándo”) y su atribución al mismo actor (dada por el factor “quién”), sin mayor codificación, marcado o interpretación del contenido más allá de la distribución en acciones, es decir, bajo una ontología basada en la contigüidad de los elementos que componen una misma acción (las cinco W), dentro de una misma sistematización informática definida por la contigüidad en la fuente y vínculos no semánticos.

20. Respectivamente: who, what, where, when, with who.

21. Latour 2005.

La naturaleza de la acción debe describirse simplemente por medio de un texto en lenguaje natural que describe un “qué”, construido por el historiador a partir de la señal, equivalente a un acontecimiento²² en la metodología histórica clásica, formulado de la forma más neutral y concisa posible, lo que la metodología analógica llamaba una “regesta”. Este formato limitado y reducido a una metodología probada garantiza la neutralidad de esta primera introducción respecto a la estructura de la base, independizada así de la interpretación posterior de los datos, garantizando flexibilidad, durabilidad y claridad. Esto tiene como objetivo principal evitar la sobrecarga cognitiva del operador, que de hecho hace imposible constituir bases incluso ligeramente complejas en la mayoría de los sistemas existentes, en los que es necesario introducir una ontología detallada desde el principio. Es inevitable, en efecto, que toda ontología detallada adquiera rápidamente un volumen y complejidad que la hacen inmanejable²³. Proponemos, por tanto, posponer el desarrollo e introducción de dicha ontología detallada a una etapa posterior del tratamiento²⁴.

D) La cuarta proposición se refiere a estas ontologías detalladas. Afirma que todas las anotaciones necesarias para la implementación de los datos contenidos en la base más allá de la ontología sumaria, a saber, codificación, agrupación, clasificación geográfica o institucional de los datos, descripción de las relaciones entre actores, entre actores y objetos, etc., se hacen posteriormente a la primera introducción; se basan en interpretaciones del material proporcionado por la primera sistematización de los datos en el momento de la primera introducción; y se superponen a los mismos sin afectar lo anteriormente cargado en la base. Esto posibilita al investigador tanteos, arrepentimientos y readaptaciones de sus propias anotaciones sin cuestionar la descripción de las acciones –el “qué”– previamente sistematizadas. Esto permite retrasar el cambio de las categorizaciones que la ontología refleja hasta un momento en el que el operador queda libre de las preocupaciones de la primera interpretación de la señal, y en el que la magnitud de los datos disponibles le ofrece una visión general del sistema estudiado que permita llegar a descripciones adaptadas al universo objeto de la investigación.

E) La quinta proposición, finalmente, es una afirmación de humildad. Por las razones que presentamos arriba, nos parece obvio que debemos renunciar a que los datos sean completamente transparentes. Los datos históricos, confusos por naturaleza y resistentes a cualquier modelo cerrado, no pueden ser objeto de una

22. No entraremos en la gigantesca literatura dedicada al concepto de acontecimiento. Sólo diremos que entendido como contenedor de acciones realizadas por un colectivo de actores en interacción los unos con los otros, o de un conjunto de acciones descritas como partes de una misma entidad por actores participantes o exteriores a ella, y considerado como bloque provisionalmente disponible para el análisis, conserva su validez.

23. Díaz-Ordóñez *et al.* 2023.

24. El listado de las clases de relaciones entre actores –de los nombres específicos que damos a cada una de esas relaciones– mencionadas hoy en día en Fichoz, por ejemplo, sobrepasa las 550 entradas. No se trata de una base de datos aparte, sino de una lista de relaciones necesaria para estructurar las relaciones que se nombren en la base de datos. Es humanamente imposible manejar un dato tan complejo al mismo tiempo que uno está repartiendo la información en acontecimientos, identificando a los actores, asignando los lugares a entidades geográficas superiores, o sea creando todas las ontologías detalladas que necesite una base.

formalización completa, absoluta. Por lo tanto, la base de datos sólo puede ofrecer datos semiestructurados, en el sentido de que todo dato contenido en la base se podrá clasificar en categorías distintas, según la problemática y los presupuestos de cada investigador, y siempre se podrá desmenuzar en elementos más finos²⁵. El sistema de tratamiento debe, en todo momento, en todas las etapas del proceso, dejar el control al operador que realizará estas sucesivas formalizaciones, en función de sus necesidades, de conformidad con las reglas desarrolladas por los fundadores de la metodología histórica.

La formalización de la ontología detallada es una actividad que requiere mucho tiempo. Acaba reportando, por otro lado, un conocimiento estrecho de la fuente y, más adelante, del fenómeno investigado. La necesidad de invertir tanto tiempo en un trabajo indispensable para sacar un pleno rendimiento del instrumento nos pareció durante mucho tiempo que podría constituir un obstáculo casi prohibitivo para la creación de las bases de datos masivas que exigía la investigación moderna. Los intentos de automatización que llevamos a cabo en un primer momento, que se basaban en las características externas del formato de los datos, no funcionaron. Sólo en documentos cuyo formato era razonablemente homogéneo se mostraron útiles, exigiendo siempre y en todo caso, verificaciones tediosas.

Parece probable que estemos en vísperas de avances decisivos en este ámbito. Varios experimentos llevados a cabo en el Instituto de Asia Oriental (IAO, Lyon, ENS) por Gunnar Knutson (este usando OpenAI) y Peter Cornwell muestran lo que la historia puede esperar de la inteligencia artificial implementada mediante herramientas masivas de aprendizaje semántico. Los primeros resultados son espectaculares. Queda por evaluar con precisión los límites de uso de estos medios y proceder a la adquisición de las técnicas necesarias. Es de notar que resulta de estos experimentos que las capacidades de las nuevas herramientas semánticas no parecen modificar en modo alguno lo dicho anteriormente respecto a la estructuración de bases de datos, sino que aceleran significativamente su implementación reduciendo, a la par, drásticamente su costo.

25. Debemos distinguir entre datos estructurados, semiestructurados y datos sin estructurar. La estructuración de los datos no tiene nada que ver con la estructura de la base. Los datos estructurados son aquellos que llevan un alto nivel de organización, lo que los hace muy fácilmente localizables en la base de datos. Por el contrario, los datos sin estructura son todo lo contrario, siendo su búsqueda y localización dentro de la base mucho más pesadas al carecer de criterios comunes para su recuperación. Mientras que la primera categoría se corresponde con un lenguaje más cercano al informático, la segunda clase es mucho más amigable para el usuario a la hora de introducir nuevos datos. Grigoriu *et al.* 2013. En el modelo de base de datos que proponemos, los datos de los campos de actor 1 y 2, relación, fecha y lugar pueden ser considerados datos estructurados; el campo de la acción y, en caso necesario, el de notas, llevarán datos semiestructurados. Los datos semiestructurados están a medio camino entre los estructurados y no estructurados. Esto permite adaptar la base a cualquier tipo de información extraída de cualquier señal manteniendo en un nivel aceptable la rapidez de búsqueda y recuperación de la información. A diferencia de lo que sucede con los datos estructurados, los datos semiestructurados permiten una introducción más sencilla por parte del usuario siguiendo unos simples patrones, fáciles de recordar, que evitan seguir complejos esquemas y criterios prestablecidos para la sistematización de la información. Este sistema, no obstante, no impide cierta uniformización de la información. Puede verse en los esquemas 03-05.

Aun así, crear una base de datos, ya sea *ex nihilo*, o agregando sus propios datos a una base ya existente, una vía que nos parece más acertada, siempre generará un coste considerable tanto en tiempo como en esfuerzos y, probablemente, en dinero. Por lo tanto, cualquier investigador, antes de decidirse a crear una base de datos para su investigación, debe plantearse la cuestión de su rentabilidad: ¿el beneficio esperado será proporcional al esfuerzo humano y financiero necesario? En muchos casos la respuesta será negativa, especialmente, si se trata de un proyecto individual. Si el proyecto es colectivo y está organizado de tal manera que los datos desarrollados por unos estén destinados a ser manejados también por otros, el coste real será entonces el de la constitución de la base dividido por el número de investigadores involucrados, un hecho que lo cambia todo²⁶.

2.2. Algunas implicaciones respecto a la organización del trabajo

Por lo tanto, la creación de una base de datos de cierto tamaño requiere la movilización o creación de un equipo. Podemos considerar esto de dos maneras. La primera consiste en reunir al mismo tiempo a varios investigadores que comparten una tarea limitada. Esta es la solución adoptada por la mayoría de los proyectos de investigación. Se corre el riesgo que desaparezcan los datos o los conocimientos necesarios para su puesta en obra cuando se deshace el equipo al terminarse el proyecto. La segunda radica en poner los datos a disposición de una comunidad, bien en forma de publicación de fuente semiestructurada, bien estructurados ya por las cinco “W” de la ontología mínima, permitiendo a otros luego retomar este conjunto para refinarlo, desarrollarlo y enriquecerlo, agregarlo de una manera que permita sumar datos adicionales. El enfoque que sugerimos, en dos capas, permite ambas estrategias. La ontología mínima de grano fino, basada en la acción, permite insertar libremente nuevas acciones atribuidas a tal o cual actor preexistente, en el formato único de las “cinco W” de nuestra ontología mínima como un elemento más en un universo granular en el que cada grano es independiente de los demás. Por ello, la introducción del nuevo dato, de la nueva acción, no desorganiza lo ya existente.

26. Lo que decimos arriba de las bases de datos parece aplicarse exclusivamente a datos cualitativos, con exclusión de los datos cuantitativos. Es un error, derivado de una equivocada oposición entre ambos conceptos generada por su historia respectiva. Consideramos un dato cuantitativo como una calidad aplicada en un contexto dado a un actor concreto, expresándose esta calidad en términos cuantitativos. Ejemplo: *Statistique générale de la France*, 1893, “Libéralités aux petits séminaires [Donativos a seminarios menores]”, 122.567 F. Transcripción en la base: [Who] Petits séminaire, Francia entera. [What] Donativos: 122.567F. [When] 1893. [With whom] Statistique Générale de la France [Where] Francia entera. Se notará que la *Statistique générale de la France* tiene la consideración de actor, ya que no sólo publicó una cantidad, sino que la elaboró previamente. Ministère du Commerce et de l’Industrie 1893. Enfocar de esta forma el dato cuantitativo como una calidad anula las fronteras materiales entre datos cuantitativos y cualitativos. Un cuadro estadístico se cargará en la base como una serie de acontecimientos, siendo cada línea del mismo un acontecimiento independiente, lo que permite, entre otras cosas, una descripción fina del contexto en el que fue producido. Cada dato puede de hecho figurar como acción independiente caracterizando a un actor cualquiera. De ser necesario, al revés, los acontecimientos extraídos de un mismo cuadro estadístico se podrán juntar de forma casi automática y exportar como un todo autosuficiente para ser sometidos a procesos estadísticos.

Lo dicho permite movilizar datos antiguos y nuevos unos al lado de los otros, porque ambos se insertan en un espacio homogéneo. Es decir, que el investigador recién llegado podrá contar con muchos datos correspondientes a los actores de su interés ya introducidos en la propia base²⁷. Este, tras insertar en la base sus propios datos bajo ontología mínima, irá aplicando a los datos que necesite, y sólo a ellos, tanto antiguos como nuevos, tanto propios como ajenos, la ontología desarrollada que habrá elaborado para su investigación sin ocuparse de lo que pudieron haber hecho sus predecesores. Esta operación de enriquecimiento era difícil en la era pre-computadora, incluso dentro de un equipo unido físicamente en un mismo lugar, fundamentalmente por la falta de un referente común potente, tal el que sí impone la máquina. Era prácticamente imposible retomar y usar las viejas “notas” de un investigador jubilado, o peor aún, de un investigador fallecido. No sólo por las propias dificultades físicas que nos pueden resultar obvias como, por ejemplo, localizar dónde se encuentran estos papeles o cómo hacernos con ellos. La falta de rigor de sus apuntes los volvía incomprensibles. Los obstáculos cognitivos, por lo menos, quedan muy reducidos por la informática. El investigador tiene que definir un modelo de “entrada” de los datos, y atenerse al mismo. So pena de perder su propio trabajo.

Las implicaciones de este enfoque, basado en lo colectivo, van mucho más allá de la mera cuestión de los datos. Al juntarlos, también junta metodologías de distintos investigadores y crea una dinámica grupal que posibilita el cruce de temas, puntos de vista y experiencias. La comunidad es, por tanto, fundamental. No se trata sólo de publicar datos totalmente estructurados (final del proceso), sino también de hacer posible el trabajo colectivo sobre datos semiestructurados bajo una ontología mínima (proceso intermedio esencial para la investigación). Esto requiere aplicaciones informáticas que ofrezcan la potencia necesaria para manejarlos. Igualmente, imprescindible resulta poner en línea las bases de datos actualmente disponibles para garantizar un libre acceso a las mismas, una retroalimentación, y la existencia de una comunidad científica realmente cohesionada. Y para ello, es necesario asegurar la permanencia del acceso colectivo a largo plazo.

Este es un problema grave, tanto más grave cuanto que el entrelazamiento de la metodología histórica y de la técnica implícito en nuestro acercamiento requiere una estrecha colaboración entre ingenieros e investigadores que desdibuja la frontera entre las dos actividades y un estricto reparto de tareas; con todos los problemas de comprensión mutua que genera el contacto de dos mundos que obedecen a disciplinas tradicionalmente consideradas opuestas.

27. Esta capacidad de la base de datos para absorber con facilidad actores nuevos requiere un sistema informático que permita la creación fácil de tales actores, la identificación sencilla de los que están en la base, así como la elemental corrección de posibles errores de identificación, todo ello a voluntad y sólo a voluntad del investigador. Cosas que las bases puramente relacionales hacen prácticamente imposibles y que en cualquier otro sistema requieren una estructuración de los datos en entradas independientes las unas de las otras, cada una dando cuenta, y plena cuenta, de una molécula de un dato completo, lo que da al manejo de la base una total flexibilidad.

El grupo Fichoz ha desarrollado procedimientos destinados a facilitar la implementación de los principios descritos anteriormente, de tal forma que los datos de fuentes homogéneas de varias clases se pueden sistematizar en las bases de datos de forma automática o semiautomática: listados de población, censos de contribuyentes, escrituras censales, registros bautismales, de matrimonio o de defunciones, genealogías, etc. Todo ello destinado a facilitar la introducción de datos y su comprensión a través de la mayor homogeneización posible. El proceso de exploración progresiva de datos permite que el fenómeno histórico estudiado se revele cada vez a niveles más profundos. Así pues, cabe reflexionar acerca de la relación entre la instrumentación, en el sentido más amplio del término, y la producción de conclusiones válidas.

3. FICHOZ: UNA HERRAMIENTA PARA EL TRATAMIENTO DE LOS DATOS

Es imposible mostrar en las siguientes páginas la infinitud de usos para los que nos aprovechamos de la base de datos Fichoz. Sin embargo, nos proponemos ilustrar los elementos clave planteados en el apartado metodológico exponiendo algunos ejemplos concretos que muestren el nivel de la granularidad al que reducimos la información que percibimos de la señal que nos entregaba la fuente. Para dichos ejemplos hemos escogido una documentación de forma intencional, simplemente, porque la conocemos bien: los papeles de la Diputación de Navarra a finales del siglo XVII. Sin embargo, el lector podrá comparar estos ejemplos con los de su propia documentación ya que, como verá, el sistema expuesto es aplicable a cualquier investigación histórica. Por ello, la documentación seleccionada solamente tiene importancia como herramienta para la demostración de la viabilidad de la base. Tal y como ya se ha dicho, el cúmulo de datos necesarios para cualquier estudio, incluso cuando este sea reducido, puede acabar superando ampliamente las capacidades cognitivas del investigador. Veamos, pues, las posibilidades que se nos ofrecen para evitarlo.

3.1. Hacia el pensamiento por “acciones”

Como lo dijimos arriba, todo documento puede reducirse a diferentes “acciones”, simplificando enormemente la dificultad presentada durante la extracción de datos y su posterior manejo y análisis. Veamos un ejemplo:

En la misma sesión se resolvió que respecto de haber cumplido Joaquín de Lete, correo, el orden del reino en ir a los veinte a la Puebla de Montalbán con las cartas para el señor conde de Oropesa, sino que fue a la Corte y se las entregó a don Pedro de Urrutia, su secretario, se le quiten de los 225 reales que estaban da[dos] por semejantes viajes cincuenta reales dándosele solamente 175 reales.

Don Miguel Jerónimo de Aranguren²⁸

28. Archivo General de Navarra, Reino, Actas Diputación, l. 5, ff. 611v-612r.

Esta información puede reducirse a una única acción siguiendo los criterios expuestos de las cinco “W”, a las que añadimos un campo de “Notas” para posibles datos extra, un campo de “Fuente”, necesario en todo documento histórico, y un campo “Relación” que resume de forma provisional en una breve etiqueta la relación entre los dos actores mencionados, para agilizar el manejo de los datos. Los tres apartados últimos no pertenecen *stricto sensu* a la ontología mínima, pero resultan útiles y en todo caso se podrán modificar posteriormente de ser necesario al proceder a la aplicación de la ontología desarrollada²⁹.

Acción A.

Actor 1 (quién realiza la acción/quién/who): NAVARRA (Diputación)

Actor 2 (quien recibe la acción/a quién/with who): LETE, Joaquín

Relación (entre Actor 1 y Actor 2): libranza

Acción (qué/what): La Diputación de Navarra libra 175 reales a Joaquín de Lete por haber entregado en la Corte a don Pedro de Urrutia las cartas para el conde de Oropesa

Fecha (when): 1691=08=01

Nota (para añadir información adicional): descuento de 50 reales por no haber ido hasta la Puebla de Montalbán como tenía ordenado

Dónde (where): Pamplona <Spain>

Fuente (de archivo o bibliográfica): *Archivo General de Navarra, Actas Diputación, l. 5/f. 614v*

Acontecimiento: 00000001

Pero debajo de esta libranza subyace otra acción: el haber entregado Lete a don Pedro de Urrutia cartas para el conde de Oropesa. Conforme al principio de atomización, tenemos que crear otra entrada para recoger el hecho:

Acción B.

Actor 1 (quién realiza la acción/quién/who): LETE, Joaquín

Actor 2 (quien recibe la acción/a quién/with whom): URRUTIA, Pedro

Relación (entre Actor 1 y Acto r2): entrega cartas

Acción (qué/what): Joaquin Lete, correo, entrega en la Corte cartas para el Conde de Oropesa a don Pedro de Urrutia.

Fecha (when): 1691-00-00

Nota (para añadir información adicional): —

Dónde (where): Madrid <Spain>

Fuente (de archivo o bibliográfica): *Archivo General de Navarra, Actas Diputación, l. 5/f. 614v*

Acontecimiento: 00000001

Debajo de la Acción B subyace más información para una acción C: Pedro de Urrutia sirve de intermediario postal entre la Diputación de Navarra y el conde de

29. En las descripciones de entradas que siguen, notamos estos tres campos en cursivas.

Oropesa. Tenemos que introducir un tercer actor, un “With who” bis, representado por el “With who” original. Lo llamaremos Actor 4³⁰.

Acción C.

Actor 1 (quién realiza la acción/quién/who): NAVARRA (Diputación)
 Actor 2 (quien recibe la acción/a quién/with whom): URRUTIA, Pedro
 Actor 4 (actor representado por el actor 2): conde de OROPESA
Relación (entre Actor 1 y Actor 2): intermediario postal
 Acción (qué/what): cartas enviadas por la Diputación de Navarra al conde de Oropesa se entregan en la Corte a Pedro de Urrutia
 Fecha (when): 1691-00-00
Nota (para añadir información adicional): —
 Dónde (where): Madrid <Spain>
Fuente (de archivo o bibliográfica): Archivo General de Navarra, Actas Diputación, l. 5/f. 614v
 Acontecimiento: 00000001
NB: la relación entre Actor 4 y Actor 2 no se explica. Por convención significa que el Actor 2 representa al Actor 4.

Se puede añadir a las tres acciones un último campo, que va a llevar un número de acontecimiento (de elección libre), que será el mismo en las tres y permitirá llamarlas juntas, como constitutivas de un mismo acontecimiento (Acontecimiento: 00000001).

Como puede verse, la información se ha desgranado hasta su nivel más bajo: la acción. Nos queda claro que la Diputación de Navarra libró una cantidad a Joaquín de Lete (Acción A) por llevar a cabo una acción (Acción B) que la Diputación le había encomendado (Acción C). Toda esa información se debe fragmentar según los criterios expresados, de tal modo que nada nos impida buscar cualquier libranza realizada por la Diputación a cualquier actor por cualquier motivo o, alternativamente, las percibidas por Joaquín Lete de cualquier institución o persona, y hacer de cada caso un caso específico con carácter y contexto propio. El proceso normal consistiría en crear la Acción A en la fase de ontología sumaria, las otras dos en la fase de ontología detallada. En esta misma a cada uno de los personajes se le atribuiría un identificador arbitrario de la forma 00000000, que permitiría localizarle independiente de cualquier otro en la base y acceder a su curso de vida. A cada lugar se le atribuirían también características propias: sus coordenadas geográficas (para hacer mapas, entre otras cosas), su rango administrativo (Pamplona: capital de Navarra, Madrid: residencia de la Corte, etc.). El apartado “Relación” se refinaría en función de una tabla general de relaciones elaborada adrede³¹, y así

30. Actor 3 sería un representativo de Actor 1. En este caso no se menciona ninguno.

31. No se trata de elaborar una base de datos aparte, sino de desarrollar una lista donde quedan registradas las relaciones a modo de etiqueta. Es decir, si para los matrimonios vamos a usar la relación-etiqueta “matrimonio”, esto debe hacerse de forma homogénea por todos los usuarios. Así, se evita que algunos introduzcan relaciones-etiquetas diferentes (nupcias, enlace, casamiento, alianza matrimonial, entre otras) que hagan que parte de los datos se pierdan a la hora de su recuperación.

consecutivamente, en función de las problemáticas abordadas. Esto se traduce en que, al preguntar a la base sobre cierto tipo de relación –una libranza, por ejemplo– nos dará todas las existentes.

Por otro lado, en esa misma libranza observamos otra información que puede resultarnos interesante: “Joaquín de Lete, correo”. Gracias a este documento se nos descubre que este personaje desempeñó el oficio de correo del reino de Navarra, información adicional que no conviene olvidar ya que aparte de iluminar la libranza presente, puede resultarnos útil en el futuro. Son muchas las ocasiones en que un documento nos reporta una información incidente, adicional a la que en principio buscábamos, lo cual puede plantear algunas dudas: ¿merece la pena almacenar dicho dato? ¿Cómo hacerlo de una manera eficaz, sin que quede olvidado? Respecto a la primera cuestión, consideramos que todo dato, sea de nuestro interés o secundario, debe ser siempre registrado³². Respondiendo acerca del cómo, Fichoz nos deja abrir una nueva entrada bajo los mismos criterios que nos permitirá su rápida recuperación en el futuro:

Acción D.

Actor 1: AUSTRIA, Carlos, España, king [Carlos II]³³

Actor 2: LETE, Joaquín

Relación: nombramiento

Acción: el rey nombra a Joaquín de Lete correo del reino de Navarra

Fecha: 1691<08<01

Esto debe hacerse con todas las relaciones que establezcamos, pues funcionan como etiquetas que nos permitirán solicitar a la base la recuperación inmediata a todos los datos que lleven dicha etiqueta.

32. Una base de datos colaborativa ofrece al investigador información de diferente clase. Por un lado, cuando otros investigadores han trabajado ya sobre algún personaje o institución que nos interesa, sus datos ya estarán registrados en la base. Así, obtendremos algunos datos que nos resulten esenciales y otros que nos proporcionen una información más adicional o complementaria. Si un investigador interesado en la historia administrativa introduce datos sobre un personaje en concreto, tendrá en cuenta, principalmente, sus nombramientos y méritos administrativos. Si otro investigador con interés en la literatura, posteriormente, se acerca al mismo personaje que, además de ocupar un cargo administrativo, escribió un texto de su interés, localizará fácilmente los datos sobre su carrera administrativa. Lo mismo sucederá a la inversa y en un sinfín de situaciones, pues cada actor será partícipe de numerosas realidades sociales. Sea cual sea nuestro objeto de estudio, son todas estas realidades sociales las que conforman la realidad verdadera del actor investigado y, en consecuencia, todas ellas se han de tener en cuenta en mayor o menor medida. Dedieu 2004, p. 7.

33. Tal y como se señalaba anteriormente, los campos de actor 1 y 2, relación, fecha y fuente contendrán datos perfectamente estructurados. Siempre que queramos introducir a un rey haremos lo siguiente: APELLIDO, Nombre, king [Nombre seguido de la ordenación en números romanos, como lo haríamos al escribirlo en un texto]. Lo mismo haremos para los nobles, cambiando “king” por su título nobiliario; con los frailes, cambiando “king” por “fray”; o con los oligarcas municipales, introduciendo el término “don”. El campo de fecha seguirá un modelo prestablecido: año=mes=día cuando sepamos la fecha exacta; año<mes<día cuando sepamos que un acontecimiento ha sucedido antes de la fecha señalada, tal y como se muestra en la acción D; o año>mes>día cuando sepamos que la acción es posterior. Esto permitirá a la base de datos ordenar cronológicamente la información sin que el usuario tenga que preocuparse por esta cuestión. La relación se escogerá de unas relaciones prestablecidas con anterioridad (matrimonio, pleito, pago, etc.). Esta estructuración de los datos, sencilla con algo de práctica, permite una gran flexibilidad de los datos y su rápida recuperación. Sobre el campo de acciones y fuente y la semiestructuración de los datos que contienen se hablará más adelante.

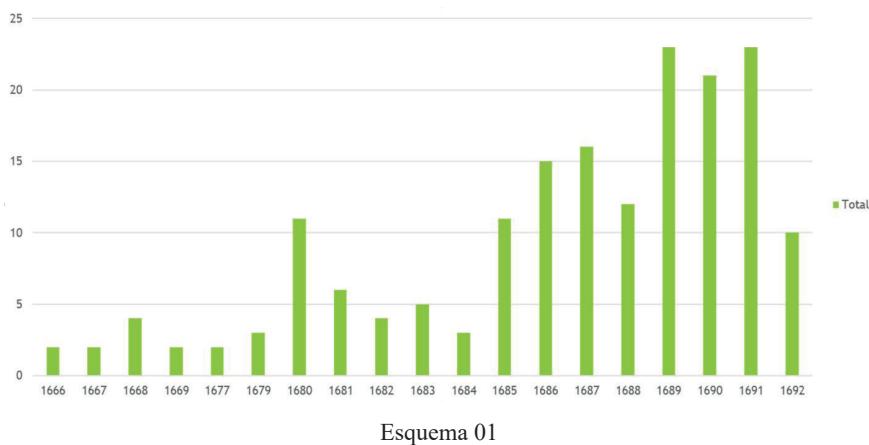
Nota: —

Fuente: Archivo General de Navarra, Actas Diputación, l. 5/f. 614v.

Como puede verse, se trata simplemente de abrir un nuevo registro con la nueva información. Cada registro, independiente del resto, no afecta en nada al resto de los datos. Sin embargo, nos permite saber que Lete era correo, que realizó una entrega de cartas al conde de Oropesa y que por dicha acción recibió una libranza de la Diputación de Navarra. Todo ello simplemente pidiéndole a la base que nos dé la información de este individuo. Cuantos más datos individuales sistematizemos, más información nos revelará la base. Por tanto, se resuelve la cuestión de cómo evitar que los datos queden aislados y olvidados, es decir, inservibles.

De un mismo documento —relativamente pequeño— hemos extraído cuatro acciones. De tres de ellas no se nos dice la fecha, por lo que el lector podrá observar como simplemente hemos utilizado los símbolos “<” o “-” para señalar que para la fecha indicada ya se había producido la acción reseñada. En el futuro, el mismo investigador u otro que encuentre una fuente donde se especifique una de las fechas que faltan podrá cambiar dicho símbolo por el “=”, que representa una fecha exacta, así como actualizar la referencia archivística o bibliográfica. Siendo el dato, del punto de vista informático, independiente de cualquier otro, el cambio no tiene incidencias ajenas. Cuando más arriba sugeríamos dejar para un momento posterior la elaboración de una ontología compleja, también hacíamos referencia a este progresivo desarrollo de la atomización. Sólo un sistema flexible nos permite situar cronológicamente un dato no de manera absoluta, sino relativa. Es decir, sabiendo, sin más, que se produce antes o después de una fecha concreta.

Un análisis exhaustivo de un fondo como el de la Diputación de Navarra para el reinado de Carlos II (1665-1700) implica la lectura y sistematización de alrededor de 2.500 páginas. No obstante, ningún buen trabajo puede conformarse con el examen de un único fondo documental, por lo que, para la conclusión del trabajo final, la documentación consultada deberá ser mucho mayor. No es exagerado afirmar que tal volumen documental necesita ser tratado mediante una base de datos elaborada según los criterios que acabamos de definir para no desbordar al investigador. Las más de mil seiscientas ochenta reuniones mantenidas por la Diputación en el reinado referido tratan temas de todo tipo, desde libranzas de dinero hasta la organización de las vísperas y misas solemnes en honor al copatrono de Navarra, san Francisco Javier, pasando por muchas otras cuestiones. En cualquier caso, la variedad de todas ellas no altera en nada la gestión que debemos hacer de dicha información. Esto es gracias al principio de homogeneización y encapsulamiento de cada dato. Por ejemplo, cuando la Diputación anota en sus libros de actas la recepción de un “papel del señor virrey en que remite la cédula de señalamiento de salarios del visitador y sus ministros”, la sistematización de la información es idéntica:



Acción E.

Actor 1: VELASCO AYALA, Antonio, Fuensalida [count]³⁴

Actor 2: NAVARRA (Diputación)

Relación: traslado

Acción: el virrey de Navarra da a la Diputación una cédula en la que se señalan los salarios del juez visitador y de sus ministros

Fecha: 1678=11=26

Nota: —

Fuente: *Archivo General de Navarra, Actas Diputación, l. 5/f. 25v-28r*

Como podemos ver, cualquier documento puede ser tratado mediante esta simple fórmula. El atomizar los datos en segmentos homogéneos proporciona un acceso rápido y eficaz a su contenido, que autoriza la creación de documentos sintéticos que permitan dominar el gran volumen de los datos. Veamos algunos ejemplos. Con un simple clic podemos recuperar todos los datos introducidos en la base referentes al Actor 1 o Actor 2, en apenas uno o dos segundos. Volvemos a incidir en el proceso de aislamiento de cada acción, lo cual hace posible que la recuperación sea tan sencilla como inmediata. Por otro lado, también es posible recuperar el número exacto de documentos de las mismas características (libranzas, asientos, contratos, órdenes, etc.). Supongamos que nos interesa saber el número de libranzas efectuadas por la Diputación entre 1666 y 1692. Aquí las tenemos ordenadas por años:

Esquema 01. Número de libranzas por años emitidas por la Diputación de Navarra. 1666-1692

El gráfico que vemos se elabora mediante la exportación de los datos introducidos en Fichoz a un fichero Excel o a una hoja de cálculo de Libre Office. Es

34. Virrey de Navarra entre 1676 y 1681.



Esquema 02.1

relativamente sencillo. Una vez introducidos los datos, su recuperación no requiere más que algunos segundos. Únicamente se debe hacer la pregunta siguiente a la base:

Actor 1: Diputación Navarra

Relación: libranza

Esta pregunta y dos segundos de espera son suficientes para que la base nos enseñe todos los registros encontrados. En este caso, 376.

Esquema 02: Libranzas emitidas por la Diputación de Navarra. 1666-1692 (formato analógico y digital)

Esquema 02.1. Formato analógico³⁵

35. El sistema de clasificación de la acción en Fichoz no supone, en sí mismo, una novedad respecto al sistema empleado por el profesor Didier Ozanam para la recogida de datos de nombramientos administrativos que comentábamos más arriba. Cada ficha, en este caso, se correspondería con una libranza realizada por la Diputación (actor 1) en Pamplona (dónde), en una fecha concreta entre 1666-1692 (cuándo), por una cantidad de dinero en pago de un servicio (qué, la acción) a quien recibe dicha cantidad (con quién). Estas fichas nos permitirían hacer el mismo gráfico que hemos obtenido gracias a la base, eso sí, de una forma mucho más lenta y tediosa. Lo que determina el modelo que proponemos para la creación de la base de datos no se basa tanto en la comprensión informática de la herramienta que utilizamos, a pesar de que ello no deje de ser importante, sino en la comprensión de las necesidades específicas de la investigación que queremos llevar a cabo y el establecimiento de un proceso eficaz para el procesamiento de la información. Dicho modelo, en su planteamiento, no difiere demasiado del modelo empleado por las fichas, aunque su tratamiento en una base de datos si mejora exponencialmente sus resultados. Fotografía del autor de un fichero donado por Jean Pierre Dedieu.

CNR5		Nbr query	Life cur	D.A. ind	D.A. corp	D.A. cult	D.A. artf	Gend/age	Act lgroup	Lg text	Action exp	Balots	Arraz un	Places	Sutes	Group	G / A/A	Feat
		Twin query	Checking	Duplicate	6001790C	WriterNet	Set chn	Alph.	Place	Lg names	Twin actors	Query PPI	Arraz m.	Deeds	Dictionary	New group	Source ar.	I ⁺ II ⁻
		NAVARRA (Diputación)	0001927C							ALEOS, Francisco				[2] La Diputación de Navarra libra 500 ducados al padre				
		NAVARRA (Diputación)	0001772C							PAYEE	00076238A	CONVENTO Santa Engracia de		La Diputación de Navarra libra 35 ducados en concepto de				1667-09-13
		NAVARRA (Diputación)	0001772C							PAYEE	0005056C	PAMPLONA (municipality)		La Diputación de Navarra libra 150 ducados a la ciudad de				1687-09-13
		NAVARRA (Diputación)	0001772C							PAYEE	0002598C			La Diputación de Navarra ordena dar libranza de los 1.552				1668-03-08
		NAVARRA (diputación)	0001772C							PAYEE	0164531A	BALANZA, Miguel		La Diputación de Navarra libra 12 ducados y una carga de trigo				1668-03-08
		NAVARRA (diputación)	0001772C							PAYEE	0157745A	INCIOA, Juana		La Diputación de Navarra libra 12 ducados a Isabel de Labayen				1668-03-08
		NAVARRA (diputación)	0001772C							PAYEE	0165102A	LABAYEN, Isabel		La Diputación de Navarra libra 100 reales a Isabel de Labayen				1668-04-01
		NAVARRA (diputación)	0001772C							PAYEE	0164920A	COMPANIA Jesús		La Diputación de Navarra libra a la Compañía de Jesús el				1668-12-05
		NAVARRA (Diputación)	0001772C							PAYEE	0164920A	PEREDA URTASUN, Diego		La Diputación de Navarra libra 100 ducados de ayuda de costa para los				1677-03-10
		NAVARRA (Diputación)	0001772C							PAYEE	0164920A	EGUA, Nicolás		La Diputación de Navarra libra 100 ducados de ayuda de costa para los				1677-03-10
		NAVARRA (Diputación)	0001772C							PAYEE	0158100A	LOPEZ CALVO, Jorge		La Diputación paga a Jorge López Calvo 57 reales por dos				1677-04-04
		NAVARRA (Diputación)	0001772C							PAYEE	0033375A	AREVALO MONTENEGRO, Alonso		La Diputación de Navarra, por orden del verry de Navarra,				1678-10-30
		NAVARRA (Diputación)	0001772C							PAYEE	0033375A	AREVALO MONTENEGRO, Alonso		La Diputación de Navarra entrega a Arevalo, visitador de los				1679-02-23

Esquema 02.2

Esquema 02.2. Formato digital: vista de Fichoz

Si nos interesase una búsqueda más refinada, podríamos acotar cronológicamente (cuándo), por individuos que reciben dichas librazas (Actor 2) o por la fuente utilizada. Como el lector apreciará de las imágenes expuestas, el aislamiento de cada dato, siempre de una forma homogénea, es lo que da forma a todo el sistema y lo hace eficaz en la recuperación de la información (tanto en su versión analógica como informática). El lector también habrá advertido un punto fundamental: la calidad gráfica de la puesta en pantalla de los datos (solo en la versión informática). Sólo en parte se debe a nuestro sistema de atomización, en gran parte al paquete de base de datos que usamos, FileMaker. Es un punto fundamental: el ojo del investigador es, en todo caso, el último eslabón de la cadena informativa. La calidad gráfica de la publicación y la facilidad de ordenar en pantalla los datos son condiciones esenciales a la hora de elegir un paquete de bases de datos con fines de investigación, es decir, un campo en que no se tiene de antemano un listado cerrado de publicaciones posibles. Esto elimina de entrada casi todos los paquetes de bases de datos existentes, empezando por los más famosos. Y desde luego las hojas de cálculo, cuya ergonomía es bastante cuestionable.

En cualquier caso, el Esquema 01 es la representación de la cuantificación de un hecho, en este caso, el número de librazas. A partir de aquello, y con el apoyo visual del propio gráfico, es posible identificar tendencias, analizar el motivo de los pagos y relacionarlos con el contexto histórico. Es decir, proceder a su análisis. La propia base ofrece la posibilidad de sistematizar todos los acontecimientos y consultar los ya registrados por investigadores anteriores. Es decir, recuperar una información muy valiosa para la contextualización de cada acción. Aunque no se pone en duda la posibilidad de realizar cuantificaciones similares empleando fondos documentales iguales o más extensos mediante otros sistemas, creemos que este es, por la estrecha integración que establece entre los datos y el gráfico, un método útil, eficaz y rápido para realizarlo.

Como se viene señalando, toda la documentación puede –y en nuestra consideración, debe– ser reducida a acciones. Cada acción nos revela un elemento nuevo más del complejo universo en el que el actor al que estudiamos tuvo que desenvolverse, y lo hace de forma cronológica.

Esquema 03

Esquema 03. Regidores del Ayuntamiento de Pamplona por burgos entre 1646 y 1843

En el esquema 03 se nos muestran los 2.212 registros contenidos en la base de datos correspondientes a los regidores de Pamplona que ejercieron su cargo entre 1646 y 1843. Como puede apreciarse, el campo de acción³⁶ muestra la información de una forma muy organizada. Cualquier individuo que ejerciese como regidor se introduce de la siguiente forma: PAMPLONA [municipality] (Actor 1); member/miembro (Relación); APELLIDO, Nombre del personaje (Actor 2); “regidor de Pamplona por el burgo de” y el barrio al que corresponde (Acción); fecha de inicio y fecha final (cuándo). Los 2.212 registros no han sido introducidos por un único investigador. Sin embargo, la fácil estrategia de sistematización de la información seguida en el campo de “acción” ha permitido una homogeneización significante de los datos introducidos por diversos usuarios. Lo mismo sucedería si en el futuro un usuario quisiera registrar a los regidores anteriores a 1646. Esto permite que al solicitar a la base que recupere a los miembros del regimiento, solo sea necesario emplear las palabras clave “regidor” y “Pamplona” para que nos revele dicha información. Siguiendo este criterio, cualquier clase de nombramiento puede introducirse en la base de datos: consejeros, alcaldes, jueces, factores, arrendadores, arrendatarios, soldados, oficiales, virreyes, obispos, cardenales, etc.

Sin embargo, el campo de “acción” no siempre puede homogeneizar toda la información con este nivel de precisión. Esto se debe a la propia naturaleza compleja de las fuentes que maneja el historiador y supone cierto problema al que el investigador se ha de enfrentar use o no una base de datos. ¿Cómo almacenar toda la información de una forma lo más homogénea posible facilitando así su acceso en el futuro a diversos investigadores? Las bases de datos no ofrecen soluciones milagrosas, pero si la posibilidad de cierto nivel de homogeneización en todos los casos. Además de la alta estructuración de los datos de los campos de “Actor” 1 y 2, “Relación” y “Fecha”, que ya de por sí ofrecen una ventaja enorme para la recuperación de los datos, se pueden seguir distintas estrategias que den cierta homogeneidad al campo de “acción”. Por ejemplo, siempre que queramos

36. El campo de acción es en el que puede leerse “regidor de Pamplona por el burgo...”.

Esquema 04

sistematizar una gracia real, la descripción de la acción irá precedida por la etiqueta “merced real”, seguida por la descripción de la acción. Además, será necesario que la descripción de la acción sea clara y concisa, de tal forma que el empleo de palabras clave permita su rápida recuperación. De forma adicional, dependiendo de la complejidad de la señal, podremos utilizar el campo de “notas” para facilitar aún más la recuperación de la información del campo de “acción”.

Esquema 04. Resultados relacionados a las fortificaciones de Pamplona a través de Notas

En el esquema 04 se muestran los 479 resultados vinculados a las fortificaciones de Pamplona registrados en la base. Para ello, solo ha sido necesario incluir la etiqueta “Fortificaciones Pamplona” en el campo de notas. Esto nos permite llamar a todos los registros relacionados con dicha cuestión, independientemente de que sean nombramientos de cargos relacionados con las fortificaciones, pagos, discusiones de las Cortes o la Diputación, informes técnicos, órdenes reales o virreinales, estado de las obras, etc. Es decir, cualquier cuestión que, teniendo relación con dicha materia, no pueda homogeneizarse hasta tal punto en el campo “acción” que permita una rápida y sencilla recuperación de los datos registrados. Otras etiquetas similares se han empleado para agrupar todas las acciones que tienen que ver con el trabajo de las Cortes o la Diputación de Navarra, la amenaza de los franceses, el pago de la media anata, la concesión de mercedes, etc. Es decir, una amplia variedad de temas que se pueden recuperar en un instante a través de este sistema de etiquetas simples.

Esquema 05. Curso de vida de Vicente Ignacio Mutiloa Andueza

Así pues, en el Esquema 05, se nos muestran las 26 acciones registradas a nombre de don Vicente Ignacio Mutiloa Andueza, las cuales no siempre se pueden homogeneizar al mismo nivel que hacíamos con los nombramientos. Sin embargo, al pedir a la base que nos revele los registros existentes de este actor, nos basta un simple vistazo para comprobar los diferentes planos en los que se vio implicado: diputado de Navarra, pleiteante en algunos litigios, comprador de una gracia al rey, etc. Por ello, sea cual sea la información que nos aporte la señal, la acción es el núcleo que debe ser desgajado del documento por el investigador pudiendo haber

Esquema 05

en muchos casos más de una acción reflejada en una misma señal. Es, sin duda, lo que nos revela las diferentes realidades de cada actor.

Además, no sólo vemos en este Esquema 05 información extraída de los fondos de la Diputación de Navarra, sino también de la sección de Tribunales Reales, de los Libros de la Cámara de Comptos, de las actas de las Cortes (Archivo General de Navarra) o de los Libros de Navarra (Archivo Histórico Nacional). Esto ilustra la capacidad de sistematización de la base, que unifica la información independientemente del tipo de documento, de su cronología o de su procedencia –puede ser archivística o bibliográfica– para ser mostrada al investigador de la forma que observamos. Esto ofrece una ergonomía indiscutible al usuario, pues solo necesita un simple vistazo para hacerse una primera idea de cada uno de los actores que le interesan. Ergonomía que viene incrementada, también, por los diferentes formatos en que se nos presentan los datos para hacerlos más accesibles a las necesidades concretas de cada investigador en cada momento, tal y como se verá en el siguiente punto.

3.2. Agrupar las acciones

Vimos arriba un caso en el que resultó útil agrupar tres acciones para formar un solo acontecimiento. Existen otros casos en que conviene que las acciones vayan agrupadas, por ejemplo, para discernir rápidamente a todos los actores participantes en un documento.

Esquema 06. Agrupamiento documental. Juramento al rey por las Cortes de Navarra (1677)

El Esquema 06 nos presenta un agrupamiento, un “grouping unit”, en el que se han agregado todos los individuos del Brazo Militar y Eclesiástico de las Cortes de Navarra que juraron lealtad a Carlos II el 10 de abril de 1677. La parte alta de

Esquema 06

la pantalla describe el documento (fecha, fuente, lugar de producción, tipo documental y regesta del contenido). La parte baja enseña las acciones que describen el acto de juramento de cada uno de los interesados. Cada una de estas acciones, formando un registro independiente en la base, aparece en la tabla de acciones en el rango cronológico que le corresponde en medio de los demás acontecimientos de la vida del interesado. La descripción del documento (parte alta) figura en una tabla aparte.

Se pueden procesar así todo tipo de documentos legales: testamentos, asientos, cédulas reales, ventas, constituciones de censos, votos en asambleas, absolutamente todo. Insertando siempre la acción por lo menos en dos series de acontecimientos: la que define el documento que se analiza, la otra el curso de la vida del actor.

Esquema 07. Tabla descriptiva de entradas de los libros de la Diputación de Navarra (c. 1665)

Cada una de estas entradas alberga información sobre cada una de las reuniones de la Diputación navarra del reinado de Carlos II de la misma manera en que se mostraba en el Esquema 06, es decir, comprendiendo a todos los actores implicados en los documentos analizados.

3.3. Otros aspectos a tener en cuenta

Ya se ha dicho que aquí no se va a poder mostrar todas las posibilidades que ofrece la base. Esta contiene también un apartado genealógico³⁷, otro geográfico

37. Como prueba del potencial de elaborar y analizar genealogías a través de Fichoz, exponemos un artículo de Álvaro Chaparro Sainz donde se analizan los orígenes sociales y medios utilizados por la familia Álava para su reproducción social. Chaparro Sainz 2012.

Encontrados 1306 / 15340											
CNRS	Nro query	G.U / acciones sh.	G.U main	G.U / List detailed	Sort chro	agrh.	Action Main	Source ar.	I'	I"	F
	G0015637	1665+0+09	La Diputación de Navarra nombra dos procuradores del reino de Navarra				Actas Diputación				Pamplona <Spain>
	G0015638	1665+0+11	Acta de la Diputación de Navarra sobre la muerte del rey, diversos memoriales y notificación de				Actas Diputación				Pamplona <Spain>
	G0015546	1665+0+19	Nombramiento por la Diputación de Navarra de José de Novar, señor de Novar, para dar el pésame				Actas de sesión				Pamplona <Spain>
	G0015569	1665+0+20	Junta de la Diputación de Navarra en la que se arrienda el estanco general del tabaco en favor de				Actas de sesión				Pamplona <Spain>
	G0015635	1665+0+30	Acta de la Diputación de Navarra en que se confirma el arrendamiento del estanco del tabaco del				Actas Diputación				Pamplona <Spain>
	G0015570	1665+1+15	Confirmación del arrendamiento del tabaco en favor de Francisco de Segura por parte de la				Actas de sesión				Pamplona <Spain>
	G0015639	1665+1+19	Acta de la Diputación de Navarra en la que se confirma el arrendamiento del tabaco en favor de				Actas Diputación				Pamplona <Spain>
	G0015571	1665+1+20	Carta a favor para don Jerónimo de Clouque otorgada en firmamento a Cortes para parte del monarca				Actas de sesión				Pamplona <Spain>
	G0015572	1665+1+23	Orden del rey a la Diputación de Navarra, legado en Madrid, sobre las órdenes de Castilla				Actas de sesión				Pamplona <Spain>
	G0015573	1665+1+23	Lectura y rendición a la catedral de José de Novar, legado en Madrid, sobre las órdenes de Castilla				Actas de sesión				Pamplona <Spain>
	G0015582	1665+1+23	La Diputación de Navarra apela a la reina y al virrey para evitar la salida de 13 000 ducados de los				Actas de sesión				Pamplona <Spain>
	G0015619	1665+1+25	Lectura en la Diputación de Navarra de una carta de José de Novar en la que da cuenta de haber				Actas Diputación				Pamplona <Spain>
	G0015620	1666+0+15	Arrendimiento del derecho de saca de lanas y artíños en favor de Juan de Olague, vecino de				Actas Diputación				Pamplona <Spain>
	G0015623	1666+0+23	La Diputación de Navarra confirma a Juan de Clouque como arrendatario de las lanas y artíños del				Actas Diputación				Pamplona <Spain>
	G0015624	1666+0+20	Orden de la Diputación de Navarra para que el padre Morel no incluya una epístola en la crónica de				Actas Diputación				Sala de la Preciosa
	G0015625	1666+0+41	Carta para Francisco Felizola, presidente de la Chancillería de Valladolid, para que ampare la				Actas Diputación				Pamplona <Spain>

Esquema 07.JPG

y un diccionario de instituciones³⁸, los tres de especial interés. Se invita al lector a comprobarlo por sí mismo, accediendo a la base o a los distintos manuales elaborados para su manejo³⁹. Sin embargo, aún nos quedan por destacar algunos aspectos señalados al principio.

La durabilidad a largo plazo de la base ofrece grandes ventajas al investigador. Se deriva de dos hechos. Quizá, la más obvia, el alojamiento de todos los datos introducidos hasta el momento y en un futuro previsible, en un servicio público nacional de almacenamiento, en este caso el sistema Huma-num (CNRS - Francia) minimizando los riesgos de que se pierda dicha información. Otra, tal vez menos evidente, la posibilidad de almacenar todas las investigaciones en una misma plataforma, lo que simplifica de forma evidente la administración de copias, soportes y ficheros. Es decir, una vía que nos permita renunciar a emplear distintos soportes para el almacenamiento de los datos, lo cual es, habitualmente, motivo de confusión para el investigador cuando tiene que recuperar los datos para su análisis. Frente a los problemas que plantea crear diferentes archivos, carpetas y documentos para el almacenamiento de datos, Fichoz ofrece una garantía excepcional: el almacenamiento de todos los datos en una única plataforma en línea⁴⁰. Consecuentemente, el investigador siempre sabe dónde están los datos recopilados, los cuales puede recuperar a golpe de clic en pocos segundos. Por ello, consideramos que cualquier base de datos debe aspirar a ofrecer un formato único

38. Conocido como DIEM (Diccionario Institucional de la España Moderna), el apartado institucional constituye una tabla separada dentro de la base de datos Fichoz. En él se describen buena parte de las instituciones de Antiguo Régimen españolas, aportando una visión panorámica de las mismas. Dedieu, 2013 p. 4.

39. Se recomienda al lector acceder a los manuales elaborados para el funcionamiento concreto de Fichoz tras la lectura de estas páginas. Estas le darán una primera noción acerca de lo que esperar de este sistema. Si considera que es un sistema útil para su trabajo, en dichos manuales encontrará la información técnica necesaria para su implementación. *Fichoz fundamentals* s. f.: <https://bases.hypotheses.org/fichoz/fichoz-fundamentals>

40. Esto es una garantía frente a la pérdida de notas manuscritas o dispositivos digitales de almacenamiento, pues la información seguirá siendo accesible desde cualquier otro dispositivo informático. Esta clase de pérdidas, por desgracia, no son algo infrecuente. Una reciente noticia en prensa nos daba cuenta de la pérdida de una tesis doctoral prácticamente acabada debido a la sustracción de la mochila del doctorando, en la cual guardaba todos sus apuntes sin disponer de copia alguna. Merello 2025.

para el almacenamiento de los datos, evitando ser selectivos o excluyentes con la información que se puede introducir.

Igualmente importante resulta la posibilidad de trabajar en equipo. Una base de datos situada en línea permite el trabajo en tiempo real de diferentes investigadores. Es decir, que todos los datos que se introducen pueden ser vistos y analizados por otros investigadores en el preciso instante en el que son registrados. De manera recíproca, también los datos que uno mismo sistematiza sirven a otros investigadores. Sin importar la institución a la que se pertenece o el territorio en el que se encuentra, el investigador tiene acceso al trabajo de sus compañeros y viceversa, lo que a nuestro entender supone una auténtica democratización de la investigación.

Finalmente, cabe destacar la continua construcción en la que se encuentra Fichoz. Si bien el proyecto comenzó en 1989, este sigue y seguirá siempre en construcción en la medida en que la investigación histórica vaya avanzando. Es un proyecto vivo bajo constante revisión, corrección y reajuste. Sin alterar los criterios que marcan la forma de trabajo, jóvenes investigadores convergen con los más veteranos para la construcción de la apasionante disciplina que supone la historia, siempre de una manera profesional y ética.

4. CONCLUSIONES

Cabe concluir recordando que nuestra propuesta metodológica no es sino una más. En ningún caso pretende ser la única. Lo que aquí se ha querido exponer es lo que, a nuestro entender, suponen los criterios más eficaces para la elaboración de una base de datos para el tratamiento de la documentación histórica, de tal modo que el análisis posterior de dichos datos sea lo más sencillo y manejable posible, sin que por el camino se pierda ninguna de las informaciones recopiladas. Cualquier otro sistema o acercamiento que consiga hacer lo que definimos arriba como necesario tiene su lugar marcado, independientemente del paquete informático seleccionado, del lenguaje informático empleado (SQL o NoSQL) o de cualquier otra característica técnica necesaria.

Y más porque lo que presentamos no es sino un tipo de base de datos entre otros posibles modelos. Lo llamamos bases “de investigación”. Insistimos en ello de forma repetitiva en el cuerpo del texto: no es un depósito de informaciones brutas, sino un lugar donde se construyen datos, y datos formalizados de forma específica, con una intervención fuerte de los operadores. Es un nivel de formalización indispensable que proporciona al investigador la materia elaborada que necesita, ya lista para su uso. El mismo, sin embargo, se edifica sobre, por lo menos, otros dos niveles de bases de datos, que responden a finalidades distintas, se estructuran de forma distinta y necesitan herramientas distintas. Un primer nivel sería el de la publicación digital de las fuentes, tal como vienen, página por página⁴¹,

41. José López Yépes describió en 1998 un proyecto pionero en España que recopiló en CD-ROM las actas de las Juntas Generales de Álava. López Yépes 1998. Esto, junto con el gran desarrollo

caracterizadas por un identificador único y descritas por ontologías bibliográficas estándares, tipo Dublin Core. Partiendo de este primer nivel, un segundo nivel tendría que entregar el texto de las fuentes no ya página por página, sino entidad semántica por entidad semántica, distinguiendo notas editoriales, textos originales, títulos, elementos de paratexto, etc. y remontando en un bloque único los fragmentos de una misma entidad dispersos entre varias páginas, identificando cada elemento con un identificador único, anotándolo con una ontología sumaria adaptada, proporcionando en la medida de lo posible un texto que la máquina podría leer, y permitiendo llamar desde este nivel las páginas correspondientes del nivel inferior. Por fin, sobre esta base, el nivel de las entidades “de investigación”, atomizadas en acciones, que a su vez permitiría un acceso directo a los dos niveles inferiores.

La implementación de un modelo como el de Fichoz soluciona también otros problemas ya planteados en este artículo. Entre ellos, las posibles restructuraciones que pueden requerir las bases de datos, tal y como señalaban Díaz-Ordoñez, Rodríguez Baena y Yun-Casalilla, al permitir la incorporación en el sistema de prácticamente toda la información extraída de cualquier fuente. También el planteado por Peter Cornwell para la salvaguarda de los datos a largo plazo. Así lo demuestra la experiencia de más de 30 años.

Finalizamos reivindicando una vez más el papel de Fichoz como foro adecuado para el trabajo en equipo. Más de 200 investigadores han trabajado con esta herramienta desde 1990 y han compartido los datos que contenía. Algo no muy habitual en nuestra disciplina. Es mucha la información que el investigador guarda en sus propios discos duros sin ponerla en conocimiento de otros profesionales y, por ende, tampoco la recibe de aquellos. Ni siquiera cuando se trata de compañeros de despacho. En no pocas ocasiones, mucha de esa información se pierde. Imaginémonos la cantidad de datos recopilados para una tesis doctoral o al final de una carrera académica. Consideramos falso de sentido que esto sea así, como si aquellos datos pertenecieran únicamente a quien los ha extraído de la fuente. Creemos, por tanto, que es hora de crear e implementar herramientas que nos hagan caminar hacia el intercambio de toda esa información, siempre desde una perspectiva recíproca que beneficie a todas las partes. Esas herramientas son, bajo nuestro punto de vista, las bases de datos colaborativas, ya sea Fichoz o cualquier otra.

5. BIBLIOGRAFÍA CITADA

Artola Renedo, Andoni; Dedieu, Jean Pierre (2011), “Venalidad en contexto. Venalidad y convenciones políticas en la España Moderna” en Francisco Andújar Castillo y María del Mar Felices de la Fuente (coords.) *El poder del dinero: Ventas de cargos y honores en el Antiguo Régimen*, Madrid, pp. 29-45.

derivado por el avance no solo de la tecnología sino también de los planteamientos metodológicos de los investigadores, muestra el gran interés que las bases de datos llevan generando durante más de dos décadas como herramientas idóneas para el análisis histórico.

- Chaparro Sainz, Álvaro (2012), “La génesis social de una familia ilustrada vasca en el siglo XVIII”, *Cuadernos de Historia Moderna*, 37, pp. 177-198. https://doi.org/10.5209/rev_chmo.2012.v37.39235
- Cornwell, Peter. (s. f.), *Report on Digital Research Infrastructures*. Westminster University. file:///C:/Users/799083/Desktop/Correcciones%20articulo%20FI-CHOZ/Leidos/digital_infrastructure_report_2022.pdf
- Dedieu, Jean Pierre (2004), “Les grandes bases de données. Une nouvelle approche de l’histoire sociale. Le système Fichoz”, *Revista da Faculdade de Letras Historia*, 5 (1), pp. 101-114.
- Dedieu, Jean Pierre (2011), “Acercarse a la «venalidad»” en Francisco Andújar Castillo y María del Mar Felices de la Fuente (coords.) *El poder del dinero: Ventas de cargos y honores en el Antiguo Régimen*, Madrid, pp. 19-28. https://shs.hal.science/halshs-00652157/file/Acerarse_venalidad.pdf
- Dedieu, Jean Pierre (2013), “Fichoz 2011. Balance de una base de datos sobre la España moderna” en Antonio Jiménez Estrella, Julián José Lozano Navarro, Francisco Sánchez-Montes González y Margarita María Birriel Salcedo (coords.) *Construyendo historia: Estudios en torno a Juan Luis Castellano*, Granada, pp. 185-200.
- Dedieu, Jean Pierre (2017), “El individuo al servicio del Estado” en María Teresa Nava Rodríguez (coord.) *De ilustrados a patriotas: Individuo y cambio histórico en la Monarquía española*, Madrid, pp. 41-66.
- Dedieu, Jean Pierre (2021), “La importancia del actor. Reflexiones sobre el porvenir de la historia social”, *Estudis: Revista de historia moderna*, 47, pp. 183-200.
- Díaz-Ordóñez, Manuel; Rodríguez Baena, Domingo Savio; Yun-Casalilla, Bartolomé (2023), “A new approach for the construction of historical databases—NoSQL Document-oriented databases: The example of AtlantoCracies”, *Digital Scholarship in the Humanities*, 38 (38), pp. 1014-1032. <https://doi.org/10.1093/lhc/fqad033>.
- Fichoz fundamentals*. (s. f.). [Hypotheses]. Recuperado 16 de abril de 2025, de <https://bases.hypotheses.org/fichoz/fichoz-fundamentals>
- Floridi, Luciano (2013), *The ethics of information*, Oxford.
- Floridi, Luciano (2019), *The Logic of Information: A Theory of Philosophy as Conceptual Design*, Oxford.
- Grigoriu, Ramona-Oana; Halcu, Ionela; Sandulescu, Virginia Cristiana; Marinescu, Mariana;& Marinescu, Viorel (2013), “Integrating of structured, semi-structured and unstructured data in natural and build environmental engineering”, *Roedunet International Conference (RoEduNet)*. <https://doi.org/10.1109/RoEduNet.2013.6511738>
- Latour, Bruno (2005), *Reassembling the Social. An Introduction to Actor-Net-work Theory*. Oxford.
- López Yépes, José (1998), “Las bases de datos históricas”, *Anales de documentación: revista de biblioteconomía y documentación*, 1, pp. 99-124.

- Malavassi Aguilar, Ana Paulina (2012), “Las bases de datos como herramienta para la investigación histórica”, *Diálogos: Revista electrónica de historia*, 13 (1), pp. 194-198. <https://doi.org/10.15517/dre.v13i1.6370>
- Merello, Patricia (2025, enero 20) “Guillermo, doctorando en Sevilla, pierde toda su tesis en un McDonald’s: «Es frustrante»”, *La voz del sur*. https://www.lavozdelsur.es/actualidad/sociedad/guillermo-doctorando-en-sevilla-pierde-toda-su-tesis-en-mcdonalds-es-frustrante_327376_102.html
- Ministère du Commerce et de l’Industrie (1893), *Statistique Générale de la France...* París.
- Ontologie (informatique) (2022) en *Wikipedia*. [https://fr.wikipedia.org/wiki/Ontologie_\(informatique\)](https://fr.wikipedia.org/wiki/Ontologie_(informatique)).
- García Molina, Jesús Joaquín; Hoyos Barceló, José Ramón; Ortín-Ibáñez, María José; Sevilla Ruiz, Diego (2023), “Una Estrategia Genérica para Migración de Bases de Datos NoSQL y Relacionales”, *Actas de las XXVII Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2023)*, pp. 1-4.
- Sánchez, Jorge (2004), “Principios sobre Bases de Datos Relacionales”, *Informe, Creative Commons*, 11, 20.
- Seignobos, Charles (1901), *La méthode historique appliquée aux sciences sociales*, Lyon.