

PROCESOS FRICCIONALES DE PUESTA EN BASES DE DATOS

Florian Jatón, Dominique Vinck

S.A.C. | « *Revue d'anthropologie des connaissances* »

2016/4 Vol. 11 | pages I à XVI

Article disponible en ligne à l'adresse :

<http://www.cairn.info/revue-anthropologie-des-connaissances-2016-4-page-I.htm>

Pour citer cet article :

Florian Jatón, Dominique Vinck, « Procesos friccionales de puesta en bases de datos
», *Revue d'anthropologie des connaissances* 2016/4 (Vol. 11), p. I-XVI.

Distribution électronique Cairn.info pour S.A.C..

© S.A.C.. Tous droits réservés pour tous pays.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

DOSSIER « LO QUE LOS DATOS HACEN HACER A LAS CHS (Y VICE-VERSA) »

PROCESOS FRICCIONALES DE PUESTA EN BASES DE DATOS

FLORIAN JATON
DOMINIQUE VINCK

Los manuscritos del Mar Muerto en el umbral de su puesta en base de datos: retos

Los manuscritos del Mar Muerto fueron descubiertos en la década de 1950 en las cuevas de la zona de Qumrán en Cisjordania. Ahora forman un corpus de unos 950 manuscritos, escritos probablemente entre finales del siglo III antes de Cristo y hasta mediados del siglo primero después de Cristo. Su estudio proporciona información valiosa sobre las fuentes de los cambios espirituales de este período crucial, en donde se entremezclan las transformaciones del judaísmo y la aparición de un cristianismo primitivo. Desde rodillos bien conservados hasta fragmentos roídos por el fuego y los hongos, estos documentos son de una calidad muy variable. Peor aún, continúan deteriorándose y algunos han perdido varios centímetros al borde de página desde los años 1950. Para contribuir a la obra colectiva de des-criptación - laboriosa y controvertida – la digitalización de los manuscritos se inició en la década de 1960. Los sistemas que se ofrecen hoy en día incluyen técnicas de luminiscencia infrarroja capaz de revelar caracteres invisibles a simple vista, herramientas de segmentación de texto, de vocalización, de constitución de diccionarios integrados con el acceso a los originales digitalizados en más o menos alta definición o propuestas de reconstitución de partes que faltan, en base a las características formales de la escritura (contornos de caracteres, espacios típicos entre caracteres o palabras, etc.). Google también ha lanzado recientemente un proyecto para escanear de nuevo todos los manuscritos en ultra alta definición en colaboración con las autoridades israelíes. Varias herramientas y sistemas para el análisis de estos manuscritos – que se ha vuelto famoso a través de la popular novela *El Código Da Vinci* – conviven de forma más o menos competitiva.

El proyecto de David Hamidovic – historiador de la Antigüedad – se inscribe dentro de linajes de herramientas explotando las propiedades píxelicas de los

caracteres de los manuscritos pero esta vez con ambición de aprendizaje sobre figuras históricas importantes pero desconocidas que son sus diferentes escribas: ¿cuáles son los proyectos editoriales inscritos dentro de los manuscritos de Qumrán? En ese momento los escribas no vuelven a copiar – el monasterio es una creación más tardía: ellos dicen, añaden, cambian y modifican de acuerdo a sensibilidades y afectos muy diversos. Si los mensajes de los textos sagrados no se ven radicalmente transformados, las maneras de producirlos – a veces – cambian. Con sensibilidades, historias y referencias diversas, los escribas de los manuscritos se pueden considerar como expresiones y participantes en los cambios confesionales de la época. Mediante la recopilación y el cruce de las huellas semánticas (palabras, giros de las frases) y formales (curvaturas y distancias) que dejaron en los manuscritos, sin duda sería posible reconstruir los perfiles de los escribas, con sus diferentes intereses. Esto implicaría el registro de datos y la formación de repertorios sobre los cuales se podría injertar algoritmos de aprendizaje más o menos supervisados.

Surge entonces la pregunta del protocolo para establecer la construcción de esta base de datos. Los manuscritos son un corpus de renombre; asociarse con informáticos activos en el campo del análisis automatizado de la escritura no debe ser algo imposible. Pero el desarrollo de modelos de aprendizaje automatizado requieren un importante trabajo preparatorio para la estructuración: ¿cómo un historiador de la Antigüedad puede entender las propuestas de TI y las irreversibilidades que construyen? ¿Cómo iniciar la construcción de esta base de datos preparatoria compuesta de caracteres sin cometer errores muy costosos para corregir? Básicamente, ¿a quién interrogar sobre las mejores maneras para la práctica de la historia de la Antigüedad y el procesamiento semiautomático de los píxeles cuando se trata de iniciar un proyecto de este tipo?

Es en el marco de un seminario de investigación sobre culturas y humanidades digitales, en donde los participantes fueron invitados para discutir los temas de sus proyectos de bases de datos digitales en las ciencias humanas y sociales (CHS) que surgió la idea de este número especial. Muchos de los proyectos de bases de datos han sido de hecho llevados a cabo en el campo de las CHS: desde los proyectos individuales de los estudiantes de doctorado hasta las grandes infraestructuras de investigación nacional, la creación de colecciones de datos digitales que se pueden alimentar, buscar y compartir, monopoliza el tiempo y la energía de muchos investigadores en CHS. También al parecer, en la mayoría de estos proyectos (si no en todos) surgen fricciones, conflictos y cuestionamientos. Los investigadores de las disciplinas involucradas son actores, portadores de visiones para el futuro de su disciplina y a veces son prescriptores controvertidos de métodos y soluciones técnicas¹. Estas

1 Consultar en particular la reseña del libro *Understanding Research Infrastructures in the Social Sciences* (Kleiner et al., 2013) por Vinck (2013). Se pone de manifiesto la utilidad de tener en cuenta las infraestructuras que permiten el trabajo científico para entenderlo. Lo que parece evidente en el estudio de la *Big Science*, con sus aceleradores de partículas (Simoulin, 2007; Boisot et al., 2011),

situaciones de trabajo constituyen a menudo *pruebas* durante las cuales se establecen irreversibilidades. Por lo tanto, ciertas problemáticas actuales parecen definirse en los proyectos de bases de datos, especialmente en relación con el estado de las CHS en el mundo académico, la pertinencia de la utilización de herramientas digitales o la credibilidad y la utilidad social de las CHS.

Desde la perspectiva de una postura sociológica – bastante común en la actualidad – tomando como punto de partida los enfrentamientos y las controversias², es sorprendente ver que estos proyectos tan publicitados de constitución de bases de datos, rara vez han sido considerados como campos de investigación. Incluso si su aspecto controvertido y su acceso a bajo costo (a menudo vienen de un laboratorio vecino, o ¡incluso su propio laboratorio!) sugieren un universo propicio para las encuestas en ambas tecnologías digitales y la práctica de investigación en CHS, muy pocos son los trabajos que toman cierta distancia para producir conocimiento sobre sus dinámicas. Algunos investigadores han hecho así públicas sus reflexiones sobre lo que está en juego en estos desarrollos (por ejemplo, Currie, 2012; Heftberger, 2012; Kirschenbaum, 2007)³, pero son tomas de posición frente a opciones y orientaciones para el uso científico de estas tecnologías. Sorprendentemente, si también se refiere a la popularidad de las recientes encuestas sobre las bases de datos – sobre todo en cuanto a los estudios sociales de la ciencia y la tecnología – estas situaciones problemáticas y heurísticas de trabajo hasta ahora se han quedado sub-invertidas.

¿Qué nos dicen? Es con el fin de entender mejor lo que se está creando en estos conflictos, que ofrecemos éste dossier temático sobre el modelaje de bases de datos digitales en las CHS. Sus contribuciones tratan, cada una a su manera, de dar cuenta de episodios controvertidos y conflictuales para documentar mejor sus resortes. ¿Qué nos pueden enseñar estos momentos delicados, de fricción (Edwards *et al.*, 2011), no sólo en el modo de existencia de estas entidades importantes que son las bases de datos digitales, sino también en las dinámicas en la investigación en CHS, cuando se trata de diseñar, construir y mantener las infraestructuras de conocimiento? Como veremos más adelante, las diferentes contribuciones dan a conocer nuevas inscripciones, habilidades y articulaciones que parecen merecer mayor consideración.

los radiotelescopios, las plataformas de nanotecnologías (Merz & Biniok, 2010; Hubert, 2013) o de genómica (Peerbaye, 2004), las redes de cooperación científica (Vinck, 1999, 2012) o la historia de las ciencias naturales con sus colecciones (Bowker 2005; Bowker & Star, 1999) no parece haber dado lugar a muchas encuestas cuando se trata de CHS.

2 Que sea de forma “clásica” utilizando las controversias como reveladores de las posiciones establecidas o al estilo de las *science studies* considerándolos como procesos performativos instituyendo lo que no pre-existía (Lemieux, 2007).

3 Por ejemplo, Heftberger sobre la automatización de los análisis de películas o Currie sobre la cartografía de las controversias en Wikipedia.

ESTUDIO SOCIAL DE LAS BASES DE DATOS

Con el vuelco digital y la multiplicación de los proyectos de humanidades digitales, este dossier indica que los estudios sociales sobre las bases de datos en CHS, están llamados a multiplicarse y vienen a unirse a los estudios sobre las infraestructuras de investigación que no son en su inicio, y no conforman hasta este momento una corriente unificada. A pesar de las diferencias teóricas notables entre estos estudios, se pueden identificar cuatro direcciones principales de análisis.

La primera es de inspiración socio-histórica y es el resultado de los recientes esfuerzos de los *Annals of the History of Computing* (Grado & Bergin, 2009; Grad, 2012, 2013). Bastante desconocidos en el panorama académico francés, esos trabajos han, sin embargo, aportado elementos cruciales para entender mejor la socio-génesis del término “base de datos” (Haigh, 2009) y sus vínculos iniciales con las arenas administrativas, militares y empresariales (Bergin & Haigh 2009; Haigh, 2011; Grier, 2012; Wade & Chamberlin, 2012; Haderle & Saracco, 2013). Si las bases de datos digitales aparecen ahora como las extensiones de colecciones de datos digitales, estos trabajos históricos pudieron así documentar sus inicios frágiles y permitieron la solidificación progresiva dentro de tecnologías dedicadas (modelo relacional, lenguaje SQL, distinción teórica entre “datos” y “puntos de vista”, etc.).

Más sociológica, la segunda dirección de exploración se ha comprometido a tratar con las bases de datos en términos de “efectos” y por lo tanto tiende a ver río abajo de su proceso de diseño. Esta dirección participa, en particular, de una lectura política inspirada en la obra de Michel Foucault sobre la vigilancia. Tiende a denunciar el poder de subyugación de los individuos y su control a través de bases de datos digitales (Lyon, 2003). De inspiración crítica, esta lectura busca revelar cómo las bases de datos requieren que las personas estén dentro de categorías con atributos circunscritos (Porter, 1995; Atten, 2013) y así intenta denunciar la asimetría entre los que diseñan las bases de datos (empresas, administraciones públicas) y los que son sus sujetos. Por lo tanto, no se consideran siempre los sujetos estrictamente como pasivos y formateados por el instrumento de control como lo demuestran algunos trabajos sobre el diseño de los sistemas informáticos (entre otros, Flichy, 2013) que reflejan cómo los actores logran jugar con las bases de datos y su poder limitante, incluso pueden volver a configurar el objeto por su uso. Este punto de vista sociológico río abajo de las bases de datos puede ir hasta encapsular una lectura simbólica y cultural (Manovich, 1999, 2012). Aquí, las bases de datos se consideran como los dispositivos característicos de una nueva forma de estudio de los medios que reflejan la sociedad actual. La interpretación global del fenómeno pone entonces más énfasis en las nuevas posibilidades de acción que las mismas bases de datos serían capaces de ofrecer.

La tercera dirección de la investigación – iniciada principalmente por los estudios sociales de la ciencia y la tecnología – se enfoca más río arriba de las

bases de datos y de las dinámicas colectivas que contribuyen a su conformación, en particular en términos de co-construcción con los estudios históricos de Yates (2005) y Gardey (2008) o los estudios sociológicos de Segrestin (2004). Vinck y Penz (2008) muestran que si el diseño de las bases de datos empresariales sigue las tendencias de gestión, estas se transforman, a medida que se traducen, en bases de datos, lo que apunta nuevas oportunidades. En el mismo sentido, las investigaciones han explorado las diferencias entre los propósitos del diseño y los usos reales. Ugettho (2013), por ejemplo, muestra que, principalmente por mecanismos enredados de co-construcción de bases de datos, las diferencias se amplían fácilmente entre las funciones imaginadas y las funciones reales. La arquitectura y las funcionalidades de las bases de datos se transforman a sí mismas según como y cuando se efectúen las discusiones y a veces hasta el punto de ya no poder ser útiles para el impulso que había generado su construcción. Del mismo modo, en el campo biomédico, Dagiral y Peerbaye (2013) muestran que las bases de datos diseñadas originalmente para compartir la información se enfrentan al problema de la interpretación de estas informaciones, terminando en luchas para asegurar una interpretación “normalizada”. Por lo tanto, es difícil inducir que los efectos de la introducción de una base de datos pueda explicarse por lo que llevan, aguas arriba, como visión del mundo o de las relaciones sociales; también hay que tener en cuenta las transformaciones realizadas durante su conformación colectiva.

Una cuarta línea de investigación – muy cercana de la tercera – trata de bases de datos digitales no solo directamente como agentamientos sociotécnicos, sino como expresiones de *procesos de categorización y de clasificación*. Porter (1995) encuesta, por ejemplo, en la formación de las convenciones sociales que servirán como puntos de apoyo a las categorías de una base de datos administrativa estadounidense. Por lo tanto, da cuenta de las dinámicas sociales, políticas y económicas, que se invitan en el diseño de las categorías: empresas, jurisdicciones que compiten, agencias federales, investigadores y científicos de la computación todos tratando de influir en la dirección de los puntos de referencia que servirán como base para la organización de los datos administrativos, lo cual termina en un acuerdo compuesto de los intereses más o menos bien representados. En el mismo sentido, el trabajo de Susan Leigh Star y Geoffrey C. Bowker (2000) muestra que las prácticas de clasificación que requieren las bases de datos también requieren prácticas de puesta en invisibilidad (Star, 1989; Bowker, 2000); un trabajo de estructuración de datos también implica un trabajo de exclusión más o menos perjudicial. Tomando el caso de los cartógrafos aficionados que tratan de identificar las instalaciones para bicicletas en una base de datos compartida, Denis y Pontille (2013) también muestran que las prácticas de desarrollo de colecciones se basan en la elección de categorías restrictivas, en clasificaciones equívocas y en habilidades visuales y descriptivas bastante precisas. En los trabajos sobre las infraestructuras de investigación en la biodiversidad, Granjou *et al.* (2014) muestran que la construcción de los contenidos moviliza nuevas colaboraciones y sistemas de gestión de las

actividades de puesta en base de datos, que los investigadores en taxonomía tienden a constituir en “proveedores de datos”. En la misma línea, Dagiral y Peerbaye (2016) muestran cómo el suministro, mantenimiento y uso de una base de datos de investigación sobre enfermedades raras, contienen aspectos sorprendentemente dobles: para los actores involucrados, la *infraestructura de conocimiento* (Karasti et al, 2016a, 2016b) es a la vez estable y representativa de las enfermedades raras, y también problemática y fuente de (re)negociación. La puesta en invisibilidad de procesos de categorización y clasificación que pasan por las tecnologías digitales parece estar respaldada por las prácticas sistemáticas de la *inversión infraestructural* (Bowker, 1994), que son recursos posibles para los actores en la situación y no solo para los investigadores en ciencias sociales.

Por diferentes e interesantes que sean, estos estudios tienen sin embargo en común el hecho que pocos de ellos trataron de las relaciones que tienen – o tenían – las bases de datos digitales con el campo heterogéneo de las CHS. Sin embargo, es a menudo bajo la apariencia de proyectos de construcción, desarrollo y mantenimiento de bases de datos digitales que operarán colaboraciones (y financiación) entre los practicantes de las CHS – historiadores, filólogos, filósofos, psicólogos, sociólogos – y las Ciencias y Tecnología de la Informática (CTI) – *data scientists*, diseñadores de interfaz gráfica de usuario o ingenieros de sistemas de información – atestigua, por ejemplo, la creciente institucionalización de las humanidades digitales (Magnien y Vinck, en prensa). ¿Qué ocurre cuando hay proyectos de desarrollo de bases de datos digitales que reúnen practicantes de las CHS y de las CTI? ¿A través de qué compromisos los (des)acuerdos logran expresarse? ¿Qué caminos prestan estas prácticas colectivas para estructurar los datos? Como veremos en este dossier temático, el hecho de pedir prestado tanto a los estudios sociales sobre la ciencia como a la investigación sobre las prácticas de clasificación, hace que nuevos arreglos se pongan en práctica, compromisos se instituyan, oportunidades nazcan y nuevos seres aparezcan. Así, a través de la creación de bases de datos digitales en CHS, es todo un medioambiente de investigación que se implementa cuya ecología merece mayor consideración.

EL CASO DE LAS BASES DE DATOS PARA LA INVESTIGACIÓN

Sin duda, las bases de datos son un fenómeno importante para la ciencia. Se podría incluso decir que hay poca ciencia sin bases de datos hechas en fichas de cartón o digitales. De hecho, los investigadores estudiaron el fenómeno y cuestionaron, en particular, la relación entre las bases de datos digitales y la actividad científica. Si seguimos Latour (1989), la actividad científica consiste en un proceso de producción de huellas e inscripciones que implica el mantenimiento de un conjunto estable de recursos; es trabajando para

recopilar documentos, uno tras otro, que *cadena de referencia* (Latour, 2012) más o menos controvertidas pueden ponerse en práctica. Pero quien dice huellas, alineaciones, compilaciones y síntesis, dice también almacenamiento, intercambio y estandarizaciones: la extensión de las redes de conocimiento no puede pasar por encima de los procesos de equipamiento (Vinck, 1999, 2011) capaces de hacer conmensurables y móviles las huellas de los eventos recogidos. La constitución de una colección de datos para apoyar las búsquedas específicas parece co-extensiva a la actividad científica y eso desde mucho antes de la llegada de las tecnologías digitales (Desrosières, 2008).

En el caso de la ciencia llamada “natural”, la dinámica de la constitución de bases de datos está bien documentada (Hilgarter, 1995, 2012; MacKenzie, 2003; Beaulieu, 2004; Bowker, 2000, 2005; Hine, 2006, Heaton & Proulx, 2012; Edwards, 2013; Meyer & Schroeder, 2015). En el caso de la genómica, Bruno Strasser (2011) muestra por ejemplo, que la ciencia que antes se presentaba como experimental, actualmente sus prácticas de investigación se ven cada vez más como las ciencias naturalistas (más antiguas y a menudo menos valoradas), precisamente debido a su reciente énfasis en la recogimiento, almacenamiento e intercambio de datos. Con el advenimiento de las tecnologías digitales y con la creación de bases de datos enormes (*Big data*), los profesionales de la genómica recopilan, clasifican, estandarizan y comparten datos que a veces pasan por delante de los experimentos específicos en su propia laboratorio. Es una nueva forma de pensar y, sobre todo, de presentar la genómica que se desarrolla como consecuencia de los nuevos requisitos exigidos por las bases de datos digitales, lo que no sucede sin resistencia y transformación. Pero ¿qué pasa con las CHS? Los estudios disponibles tratan casi exclusivamente de las ciencias “duras” y muy poco se ha hecho sobre la relación entre las CHS y las bases de datos (Vinck, 2013).

Las colecciones de datos digitales en las ciencias humanas y sociales

Las bases de datos también son numerosas e importantes para las ciencias humanas y sociales. Ya sea que se limiten a proyectos individuales, que estén incrustadas en instituciones⁴ que garanticen su relativa permanencia, atadas a comunidades de investigación específicas, a infraestructuras técnicas o a normas reglamentarias (por ejemplo, de la protección de la intimidad), estas bases de datos tienen historias que serían tan útiles para rastrear y analizar. Pensamos por ejemplo, al proceso de constitución de archivos privados en los Estados Unidos en la década de 1940 para la investigación sobre la opinión pública, el desarrollo de cuantificación en CHS (en lingüística, historia y sociología

4 Por ejemplo, el Consorcio Europeo para las Infraestructuras de Investigación (ERIC), el Consejo de Archivos Europeos de Datos de Ciencias Sociales (CESSDA), el *International Federation of Data Organisations* (IFDO) o la infraestructura Huma-Num de apoyo a las ciencias humanas en Francia.

en particular), el desarrollo de los servicios nacionales de estadística o de diccionarios de idiomas, hasta los recientes desarrollos de las humanidades digitales, de colecciones de huellas digitales y del *big data*: todos son elementos que están hoy en el centro de nuestras prácticas y esta dataficación (Bastin & Francony, en este número) es favorable para nuestras propias operaciones de investigación a través de bases de datos. La estructuración e intercambio de datos también están relacionadas con reflexiones metodológicas disciplinarias (armonización de protocolos, comparabilidad, interoperabilidad, gestión de la calidad), así como con los acuerdos de equivalencia entre objetos, personas, relaciones o acontecimientos que los investigadores y la sociedad construyen (Desrosières 2008), en espera de la redefinición de la forma en su estatus. La constitución de bases de datos abre oportunidades para el tratamiento cuantitativo que depende tanto de las operaciones de clasificación que de distinciones y equivalencias establecidas (por ejemplo, tratar a todos los seres humanos como iguales, ya sean trabajadores o propietarios) que puede tener implicaciones políticas cuando los enunciados producidos por la investigación se utilizan para la decisión o la comunicación política. Una sociología de las bases de datos en CHS podría contribuir a esclarecer los procesos sociales, políticos y técnicos de la producción de conocimientos en CHS.

La constitución de bases de datos en CHS también requiere de un trabajo a menudo invisible, y de muchas operaciones, teniendo en cuenta las mejores prácticas (*best practices*) y los estándares internacionales (por ejemplo, normas internacionales de archivado de los investigadores (DDI) o aquellas que conciernen los archiveros (SADI)), para recopilar datos, prepararlos (por ejemplo, su anonimato en que se refieren a las personas que viven), armonizarlos, limpiarlos, indexarlos y documentar las condiciones de su producción (Beltrame & Jungen, 2013), preservarlos, relacionarlos y proporcionarles acceso o permitir su reutilización para responder a diferentes preguntas.

El éxito de estas bases de datos proviene en particular de su transparencia, es decir, la invisibilidad y el enmascaramiento de las operaciones de que dependen y de la hormiguita que las lleva a cabo (ver el dossier temático sobre esta cuestión en la *Revue d'Anthropologie des Connaissances* – Denis & Pontille (2012), en particular el artículo de Millerand (2012)). El usuario-investigador a menudo anhela el acceso “inmediato” (Star & Strauss, 1999) a su objeto de estudio o a los datos. No quiere molestarse con los problemas de diseño, gestión, seguridad y mantenimiento. Además, estas bases de datos parecen dar más satisfacción que transparencia al usuario.

Todo esto no está exento de dificultades en la cooperación con los expertos en ciencia y tecnología de la información, ya sea en su concepción o en su uso. Dígase lo que se diga, en discusiones colectivas aguas arriba la construcción de bases de datos en CHS siguen siendo poco equipadas. La elección del sistema de base de datos a menudo se confía a un informático del grupo de investigación o incluso constituye únicamente una opción por carencia, ya que rápidamente crea el carácter irreversible y no asegura la sostenibilidad de la base de datos.

Algunos temas de las colecciones de datos digitales en CHS

Las dificultades y las tensiones en el diseño y gestión de bases de datos de investigación en CHS permanecen sin explorar, a menudo ahogadas por el peso de las presiones de los resultados del trabajo de visibilidad: demostraciones de la herramienta y publicaciones que obedecen a menudo – y esto es comprensible – al amparo por los líderes de proyecto. Pero la oportunidad es hermosa: documentando lo que sucede en estos proyectos turbulentos, ¿no podríamos lograr generar nuevos conocimientos sobre las prácticas de investigación en CHS con las colecciones de datos digitales? ¿El aumento de la financiación en relación con la creación de bases de datos digitales en CHS no podría dar lugar a un aumento en los estudios sociales de la nueva dinámica dentro de las CHS frente a estas tecnologías? ¿Las fricciones generadas durante estos proyectos, a menudo interdisciplinarias, no serían también oportunidades para documentar mejor algunas prácticas de clasificación hasta entonces mal definidas, pero que podrían afectarnos directamente? Explorando situaciones problemáticas, a veces conflictivas, el objetivo de este dossier es repoblar las bases de datos para comprender mejor lo que está en juego con las infraestructuras de investigación en CHS. Una investigación preliminar, inédita, de Pierre-Nicolas Oberhauser (Jaton, 2014) con los investigadores en CHS, ya había puesto de manifiesto los elementos siguientes.

La problemática del compartir los datos digitales

Con el advenimiento de la tecnología digital, los datos se hacen fácilmente compartibles. Hasta el momento en que estuvieron confinados en carpetas y fichas, los investigadores (historiadores, lingüistas, psicólogos) debían hacer esfuerzos concretos para compartir sus datos con sus colegas. Como ya habían pasado mucho tiempo para producir su ordenamiento, a menudo no veían la utilidad del trabajo adicional necesario para dejar sus datos a disposición. Con la digitalización, datos y formatos se duplican con mayor facilidad. Los esfuerzos para compartir parecen reducirse; la manera de expresar los datos los haría así mismo compartibles y reduciría drásticamente el costo de su circulación.

Estas oportunidades para compartir datos digitales no se produce únicamente río abajo de su producción; también se dan aguas arriba, en términos de la colaboración para su producción. De este modo surgen otros cuestionamientos y problemas tales como la determinación de la propiedad de los datos: ¿a quién pertenecen los datos? ¿Quién tiene derecho o no a trabajarlos? Otros problemas se refieren a la producción real de estos datos: ¿quién es responsable de producirlos? ¿No son a menudo los estudiantes de doctorado? Si es así, ¿cómo éstas tareas menores compiten con la formación de trayectorias intelectuales propias?

La problemática de la colectivización del trabajo científico

Las colaboraciones científicas sugeridas por las posibilidades de compartir datos digitales introducen nuevos cuestionamientos de organización. El primero está relacionado con un « funcionamiento en circuito cerrado » (problema de *bootstrapping*) (Bowker *et al.*, 2010). Según Geoffrey Bowker, para que los datos se consideren adecuados, se deben cruzar con los datos producidos por otros investigadores sobre los mismos objetos. A partir de allí, la colectivización del trabajo científico implicará el establecimiento de acuerdos para definir los objetos capaces de colocarse en la intersección de varias preocupaciones. La cuestión es de definir a base de qué criterios se establecen acuerdos sobre estos “objetos-frontera”: ¿Cuáles son las capacidades de asociación de estos objetos-frontera y que está involucrado en sus definiciones? (Star & Griesemer, 1989; Star, 2010)

Un segundo problema relacionado con la colectivización del trabajo científico tiene que ver con la normalización de los datos y en lo concreto de las reglas de formato de datos. ¿Hasta qué punto estas reglas son únicas? ¿Cómo se producen? ¿Cómo están puestas en acción por los investigadores involucrados?

La problemática de los datos producidos

Una vez producidos – es decir, puestos en colecciones de tal forma que permitan responder a las búsquedas – los datos están generalmente sujetos a ensayos epistemológicos que ponen a prueba la solidez de la relación entre los datos y los objetos de estudio. Pero igual de importante, los datos también están sujetos a pruebas técnicas que exploran su capacidad para adaptarse a las necesidades prácticas (y cambiantes) de los investigadores. La robustez y la flexibilidad de la colección de datos: es el vínculo sutil entre estos dos atributos que parece definir una buena colección de datos digitales. A partir de ahí surge una serie de preguntas: ¿estos dos atributos logran ser tratados de forma simétrica en el proceso de producción de la colección de datos digitales? ¿Cómo actuar frente a ésta necesidad de articulación entre robustez y flexibilidad dentro de un universo colectivo y distribuido? ¿Estos dos requisitos, inducen una nueva forma de trabajar con la ingeniería informática? Si es así, ¿cuál?

FRICCIONES QUE HACEN HACER

Se entenderá, incluso si se han reunido elementos significativos de análisis sobre la performatividad de las bases de datos (Lyon, 2003; Manovich, 1999, 2012; Porter, 1996), no sólo se trata aquí de los efectos de las bases de datos sobre las prácticas en CHS⁵. Se trata también de investigar lo que sucede cuando los

5 Ni tampoco al contrario, de investigar los efectos de las CHS en, por ejemplo, la arquitectura de las bases de datos digitales.

proyectos académicos en CHS se involucran en el proceso de construcción de bases de datos digitales. A la rigidez de los “efectos en”, este dossier temático prefiere la flexibilidad de las “relaciones entre” que se establecen en situaciones de trabajo; a la voz activa de la maestría o pasiva de la dominación, favorece la voz media del “hacer hacer” (Latour, 2000). Sobre la base de la sociología de las pruebas (Akrich, Callon, & Latour, 2006; Boltanski & Thévenot, 1991), este dossier temático se inserta en aquellos momentos de fracturas y disputas – que aquí llamamos “fricción” siguiendo a Edwards *et al.* (2011) – para documentar lo que aparece; que despiden las chispas que surgen en el proceso de fricción de la configuración de bases de datos digitales en proyectos de CHS?

Cada contribución responde a esta pregunta de una manera singular. Para Gilles Bastin y Jean-Marc Francony quienes intentan pasar *inscripciones* almacenadas en servidores web (en el caso de la red social profesional LinkedIn) a *datos* que se pueden cuestionar como parte de una investigación sociológica, el seguimiento de las fricciones traza una obligación para aumentar la visibilidad: el proceso de transformación problemática que hace pasar las inscripciones de la web a datos de investigación parece depender de un ejercicio de composición entre intereses, actores y definiciones diferenciados. Una investigación basada en una base de datos digital como la que hicieron, no puede, por tanto, prescindir de una explicación de los cambios que ésta base de datos se ha hecho necesaria. Para ir desde las inscripciones hasta los datos, se debe pasar por un proceso discordante de *datafización* (Bastin & Francony, en este número), es decir de la producción de materiales a partir de los cuales se despliegan el procesamiento y los análisis estadísticos. Esta contribución muestra el ensamblaje de definiciones contradictorias de datos técnicamente y socialmente aceptables por todas las partes interesadas en la investigación (incluyendo los que se invitaron tardíamente) para constituir una base de datos.

Sophie Duchesne y Mathieu Brugidou retroceden en el camino de un proyecto de desarrollo de base de datos de encuestas cualitativas de habla francesa. Al explorar los muchos puntos de fricción – a menudo dolorosos – de este amplio proyecto, los autores hacen surgir una necesidad delicada de proyección: mientras el proyecto está todavía en su infancia éste parece pedir a sus actores que simulen – y hagan frente a – lo que podría permitir la base. A continuación definen y confrontan visiones a veces contradictorias de la base como archivo o como una herramienta de investigación, pensada en términos de operación o en términos de uso. Se muestra cómo el diseño de una base de datos se enfrenta a la inflación de las promesas de excelencia (especialmente sus efectos sobre la normalización de las prácticas de investigación) y plantea debates sobre las buenas prácticas de reutilización de datos de la investigación. El artículo documenta la dinámica y la pluralidad de enfoques y perspectivas institucionales y profesionales que dan forma a las herramientas de investigación en ciencias sociales. Hace también hincapié en la necesidad de considerar los efectos políticos de esta conformación y cuestiona a los bancos de encuestas existentes.

Por último, explorando muy precisamente los ajustes necesarios para la consecución de un proyecto de bases de datos digital en psicología, Pierre-Nicolas Oberhauser hace visibles las habilidades que tienen que construir los actores – casi a vuelo de mosca! – para provocar las relaciones entre los diferentes campos de especialización y su involucramiento. La fricción en un proyecto y los datos, no aparecen como dados; sigue siendo necesario que puedan suceder, lo que requiere un cierto trabajo creativo y de vigilancia de parte de los investigadores.

El despliegue de la red de participantes en la encuesta de Gilles Bastin y Jean-Marc Francony; la lectura institucional y de la experiencia profesional por Sophie Duchesne y Mathieu Brugidou; el anclaje situacional de las relaciones interdisciplinarias por Pierre-Nicolas Oberhauser: tres contribuciones que exploran los procesos de fricción en el establecimiento de bases de datos de formas diferenciadas. ¿Son solamente conmensurables hasta el punto de proporcionar una propuesta generalizada y reutilizable? Obviamente, el lector debe decidir. Sin embargo, creemos – y es por lo tanto bastante exploratorio – que éstas contribuciones no se limitan a explorar situaciones hasta ahora poco documentadas; sino que también ayudan a hacer que la figura “base de datos digitales en CHS” sea un mejor tema de estudio, ya que mejor equipado. De hecho, uniendo las bases de datos digitales los deseos de composición, de simulación y las habilidades situacionales, no solo se proporciona un soporte conceptual que se puede utilizar para abrir e iniciar nuevas investigaciones, sino que también se propicia la búsqueda de nuevos proyectos de creación de bases de datos.

Agradecimiento

La publicación de este dossier temático ha recibido el apoyo del Fonds des publications de la Universidad de Lausanne (Unil). Agradecemos a los colegas que han participado en los seminarios de LaDHUL (laboratorio de culturas y humanidades digitales de Lausana) y en el seminario celebrado en Lausana en octubre del año 2015 que reunía a los autores seleccionados sobre la base de un resumen extendido. Agradecemos también al Instituto de Ciencias Sociales de la Universidad de Lausana que proporciona su apoyo. Finalmente, agradecemos a Marc Barbier quién, a nombre del Comité Editorial de la RAC, se ha involucrado mucho en el acompañamiento del dossier temático, y a Guillermo Uribe por la revisión de esta traducción.

REFERENCIAS

- Akrich, M., Callon, M., & Latour B. (2006). *Sociologie de la traduction : textes fondateurs*. Paris : Presses de l'École des Mines.
- Atten, M. (2013). Ce que les bases de données font à la vie privée. L'émergence d'un problème public dans l'Amérique des années 1960. *Réseaux*, 178-179, 23-53. doi:10.3917/res.178.0021
- Beaulieu, A. (2004). From brainbank to database: the informational turn in the study of the brain, *Studies in History and Philosophy of Science. Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 35(2), 367-390.
- Beltrame, N.T. & Jungen, C. (2013). Cataloguer, indexer, encoder. Ou comment les données prennent vie, *Revue d'anthropologie des connaissances*, 7(4), 747-759. doi:10.3917/rac.021.0747
- Bergin, T.J. & Haigh, T. (2009). The Commercialization of Database Management Systems, 1969-1983. *IEEE Annals of the History of Computing*, 31(4), 26-41.
- Boisot, M., Nordberg, M., Yami, S., & Nicquevert, B. (2011). *Collisions and collaboration. The organization of learning in the Atlas experiment at the LHC*. Oxford: Oxford University Press.
- Boltanski, L. & Thévenot, L. (1991). *De la justification : les économies de la grandeur*. Paris : Gallimard.
- Bowker, G.C. (1994). Information Mythology and Infrastructure, in L. Bud-Friedman (ed.), *Information Acumen: the Understanding and Use of Knowledge in Modern Business*, London: Routledge, 231-247.
- Bowker, G.C. (2000). Biodiversity Datadiversity, *Social Studies of Science*, 30(5), 643-683.
- Bowker, G.C. (2005). *Memory practices in the sciences*. Cambridge, Mass., London: MIT Press.
- Bowker, G.C. & Star, S.L. (2000). *Sorting things out. Classification and its consequences*. Cambridge, Mass.: MIT Press.
- Bowker, G.C., Baker, K., Millerand, F., & Ribes, D. (2010). Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment, in J. Hunsinger, L. Klastrup, & M. Allen (eds.), *International Handbook of Internet Research* (pp. 97-117). Dordrecht: Springer.
- Currie, M. (2012). The Feminist Critique: Mapping Controversy in Wikipedia, in D. Berry (ed.), *Understanding Digital Humanities* (pp. 224-248). Houndmills: Palgrave Macmillan.
- Dagiral, E. & Peerbaye, A. (2013). Voir pour savoir. Concevoir et partager des « vues » à travers une base de données biomédicales, *Réseaux*, 178-179, 162-196. doi:10.3917/res.178.0163.
- Dagiral, E. & Peerbaye, A. (2016). Making Knowledge in Boundary Infrastructures: Inside and Beyond a Database for Rare Diseases, *Science & Technology Studies*, 29(2), 44-61.
- Denis, J. & Pontille, D. (2012). Travailleurs de l'écrit, matières de l'information, *Revue d'anthropologie des connaissances*, 6(1), 1-20. doi:10.3917/rac.015.0001
- Denis, J. & Pontille, D. (2013). Une infrastructure évasive : aménagements cyclables et troubles de la description dans OpenStreetMap, *Réseaux*, 178, 91-125.
- Desrosières, A. (2008). *Pour une sociologie historique de la quantification*. Paris : La Découverte.
- Edwards, P.N., Mayernik, M.S., Batcheller, A.L., Bowker, G.C., Borgman, C.L. (2011). Science Friction: Data, Metadata, and Collaboration. *Social Studies of Science*, 41(5), 667-690. doi:10.1177/0306312711413314
- Edwards, P.N. (2013). *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*, Cambridge, Mass.: The MIT Press.
- Flichy, P. (2013). Rendre visible l'information. Une analyse sociotechnique du traitement des données. *Réseaux*, 178-179, 55-89.
- Flichy, P. & Parasie, S. (2013). Sociologie des bases de données : présentation, *Réseaux*, 178-179, 9-19. doi:10.3917/res.178.0009
- Gardey, D. (2008). *Écrire, calculer, classer. Comment une révolution de papier a transformé les sociétés contemporaines (1800-1940)*. Paris : La Découverte.

- Grad, B. (2012). Relational Database Management Systems: The Formative Years. *IEEE Annals of the History of Computing*, 34(4), 7-8.
- Grad, B. (2013). Relational Database Management Systems: The Business Explosion. *IEEE Annals of the History of Computing*, 35(2), 8-9.
- Grad B. & Bergin, T.J. (2009). History of Database Management Systems. *IEEE Annals of the History of Computing*, 31(4), 3-5.
- Granjou, C., Mauz, I., Barbier, M., & Breucker, P. (2014). Making taxonomy environmentally relevant. Insights from an all taxa biodiversity inventory. *Environmental Science & Policy*, 38, 254-262.
- Grier, D.A. (2012). The Relational Database and the Concept of the Information System. *IEEE Annals of the History of Computing*, 34(4), 9-17.
- Haderle, D.J. & Saracco, C.M. (2013). The History and Growth of IBM's DB2. *IEEE Annals of the History of Computing*, 35(2), 54-66.
- Haigh, T. (2009). How Data Got Its Base: Information Storage Software in the 1950s and 1960s. *IEEE Annals of the History of Computing*, 31(4), 6-25.
- Haigh, T. (2011). Charles W. Bachman: Database Software Pioneer. *IEEE Annals of the History of Computing*, 33(4), 70-80.
- Heaton, L. & Proulx, S. (2012). La construction locale d'une base transnationale de données en botanique. Une mise en lumière du travail des « petites mains », *Revue d'anthropologie des connaissances*, 6(1), 141-162. doi:10.3917/rac.015.0179.
- Heftberger, A. (2012). Do Computers Dream of Cinema? Film Data for Computer Analysis and Visualisation, in D. Berry (ed.), *Understanding Digital Humanities* (pp. 210-223). Houndmills: Palgrave Macmillan.
- Hilgartner, S. (1995). Biomolecular Databases: New Communication Regimes for Biology?, *Science Communication*, 17(2), 240-263.
- Hilgartner, S. (2012). Selective flows of knowledge in technoscientific interaction: information control in genome research, *The British Journal for the History of Science*, 45(2), 267-280.
- Hine, C. (2006). Databases as scientific instruments and their role in the ordering of scientific work. *Social Studies of Science*, 36(2), 269-298.
- Hubert, M. (2013). *Partager des expériences de laboratoire. La recherche à l'épreuve des réorganisations*. Paris : EAC.
- Jaton, F. (2014). Approches sociologiques des bases de données, miméo, compte rendu de l'exposé de P.N. Oberhauser et D. Vinck lors du séminaire du LaDHUL du 31 mars 2014, 4-5.
- Karasti, H., Millerand, F., Hine, C.M., & Bowker, G.C. (2016). Knowledge infrastructures: Part I. *Science & Technology Studies*, 29(1), 2-10.
- Karasti, H., Millerand, F., Hine, C.M., & Bowker, G.C. (2016). Knowledge infrastructures: Part II. *Science & Technology Studies*, 29(2), 2-6.
- Kirschenbaum, M.G. (2007). *The remaking of reading: Data mining and the digital humanities*. The National Science Foundation Symposium on Next Generation of Data Mining and Cyber-Enabled Discovery for Innovation, Baltimore, MD.
- Kleiner, B., Renschler, I., Wernli, B., Farago, P., & Joye, D. (eds.) (2013). *Understanding Research Infrastructures in the Social Sciences*. Zurich: Seismo.
- Latour, B. (1989). *La science en action. Introduction à la sociologie des sciences*. Paris : La Découverte.
- Latour, B. (2000). Fracture/fractures. De la notion de réseau à celle d'attachement. In A. Micoud & M. Peroni (dir.). *Ce qui nous relie* (pp. 189-208). La Tour d'Aigues : Éditions de l'Aube.
- Latour, B. (2012). *Enquête sur les modes d'existence : une anthropologie des Modernes*. Paris : La Découverte.
- Lemieux, C. (2007). À quoi sert l'analyse des controverses ? *Mil neuf cent. Revue d'histoire intellectuelle*, 25(1), 191-212.

- Lyon, D. (ed.) (2003). *Surveillance as social sorting: privacy, risk, and digital discrimination*. London: Routledge.
- MacKenzie, A. (2003). Bringing sequences to life: how bioinformatics corporealizes sequence data, *New Genetics and Society*, 22(3), 315-332.
- Magnien, Y. & Vinck, D. (à paraître). Des résistances à la numérisation des données : les aventures du Big data au CNRS. In I. Sainsaulieu & A. Saint-Martin (dir.). *Travail, sciences et technologies : les formes contemporaines de l'innovation*.
- Manovich, L. (1999). Database as Symbolic Form, *Convergence*, 5(2), 80-99. doi:10.1177/135485659900500206.
- Manovich, L. (2012). How to Compare One Million Images? In *Understanding Digital Humanities*, Palgrave Macmillan, 249-278.
- Merz, M. & Biniok, P. (2010). How Technological Platforms Reconfigure Science-Industry Relations: The Case of Micro- and Nanotechnology, *Minerva*, 48, 105-124.
- Meyer, E.T. & Schroeder, R. (2015). *Knowledge Machines: Digital Transformations of the Sciences and Humanities*. Cambridge, Mass.: MIT Press.
- Millerand, F. (2012). La science en réseau. Les gestionnaires d'information « invisibles » dans la production d'une base de données scientifiques, *Revue d'anthropologie des connaissances*, 6(1), 163-190.
- Parasie, S. (2013). Des machines à scandale. Élément pour une sociologie des bases de données, *Réseaux*, 178-179, 127-161. doi:10.3917/res.178.0127
- Peerbaye, A. (2004). *La construction de l'espace génomique en France : la place des dispositifs instrumentaux*. Thèse de doctorat, Cachan, École Normale Supérieure de Cachan.
- Porter, T.M. (1996). *Trust in Numbers*. Reprint Edition. Princeton, N.J.: Princeton University Press.
- Segrestin, D. (2004). *Les chantiers du manager*. Paris : Armand Colin.
- Simoulin, V. (2007). Une communauté instrumentale divisée... et réunie par son instrument, *Revue d'anthropologie des connaissances*, 1(2), 221-241.
- Star, S.L. (1989). *Regions of the Mind: Brain Research and the Quest for Scientific Certainty*. Stanford: Stanford University Press.
- Star, S.L. (2010). Ceci n'est pas un objet-frontière ! Réflexions sur l'origine d'un concept, *Revue d'anthropologie des connaissances*, 4(1), 18-35. doi:10.3917/rac.009.0018
- Star, S.L. & Griesemer, J. (1989). Institutional ecology, 'Translations' and Boundary objects: amateurs and professionals on Berkeley's museum of vertebrate zoology. *Social Studies of Science*, 19(3), 387-420.
- Star, S.L. & Strauss, A. (1999). Layers of silence, arenas of voice: the ecology of visible and invisible work, *Computer Supported Cooperative Work*, 8, 9-30.
- Strasser, B. (2011). The Experimenter's Museum: GenBank and the moral economies of biomedicine, *ISIS*, 102(1), 60-96.
- Ughetto, P. (2013). Utiliser une base de données en organisation. La recherche de l'instrument, *Réseaux*, 178-179, 197-222. doi:10.3917/res.178.0197
- Vinck, D. (1999). Les objets intermédiaires dans les réseaux de coopération scientifique. Contribution à la prise en compte des objets dans les dynamiques sociales, *Revue française de sociologie*, 40(2), 385-414.
- Vinck, D. (2011). Taking intermediary objects and equipping work into account when studying engineering practices, *Engineering Studies*, 3(1), 25-44.
- Vinck, D. (2013). Pour une réflexion sur les infrastructures de recherche en sciences sociales, *Revue d'anthropologie des connaissances*, 7(4), 993-1001.
- Vinck, D. & Penz, B. (dir.) (2008). *L'équipement de l'organisation industrielle. Les ERP à l'usage*. Paris : Hermès.
- Wade, B.W. & Chamberlin, D.D. (2012). IBM Relational Database Systems: The Early Years. *IEEE Annals of the History of Computing*, 34(4), 38-48.
- Yates, J.A. (2005). *Structuring the Information Age: Life Insurance and Technology in the Twentieth Century*. The Johns Hopkins University Press.

Florian JATON es estudiante de doctorado en sociología – estudios sociales de ciencia y técnica en la Universidad de Lausana (UNIL). Es miembro del Instituto de Ciencias Sociales de la UNIL, del STS Lab et del Laboratorio de culturas y humanidades digitales (LaDHUL). Se encuentra en estancia de investigación en la Universidad de California a Irvine, en el laboratorio EVOKE. Su tesis esta basada en un trabajo etnográfico en un laboratorio de informática especializado en el tratamiento de la señal digital. Su estudio de laboratorio se orienta en la constitución de referenciales de medida para el trabajo algorítmico, en las prácticas de programación informática y en la conformación de modelos informáticos.

Dirección	Université de Lausanne (UNIL) Institut des Sciences Sociales – LaDHUL Quartier Mouline - Géopolis CH-1015 Lausanne (Suisse)
Correo	florian.jaton@unil.ch, fjaton@uci.edu

Dominique VINCK es Profesor ordinario en la Universidad de Lausana (UNIL) y docente en el Collège des Humanités de la Escuela Politecnica Federal de Lausana (EPFL). Miembro del Instituto de Ciencias Sociales de la UNIL y del STS Lab, dirige el Laboratorio de culturas y humanidades digitales (LaDHUL). Sus investigaciones conciernen la sociología de la ciencia y de la innovación. Trabaja actualmente en el campo de la ingeniería de las culturas y humanidades digitales. Publicó entre otros: *Engenhieros no Cotidiano. Etnografia da Atividade de Projeto e Inovação* (Fabrefactum Editora, Belo Horizonte (Brasil), 2013), *Pratiques de l'interdisciplinarité* (PUG, Grenoble, 2000), *Ciencias y sociedad. Sociología del trabajo científico* (Gedisa, Barcelona, 2015), *L'équipement de l'organisation industrielle. Les ERP à l'usage* (Hermes, 2008), *Les nanotechnologies* (Le Cavalier Bleu, 2009), *Comment les acteurs s'arrangent avec l'incertitude* (EAC, 2009), *Les Masques de la convergence* (EAC, 2012), *Ingénieur aujourd'hui* (PPUR, 2015), *Sciences et technologies émergentes : pourquoi tant de promesses ?* (Hermann, 2015), *Humanités numériques : la culture face aux nouvelles technologies* (Le Cavalier Bleu, 2016).

Dirección	Université de Lausanne (UNIL) Institut des Sciences Sociales – LaDHUL Quartier Mouline - Géopolis CH-1015 Lausanne (Suisse)
Correo	dominique.vinck@unil.ch
