

“A la preservación de datos... ¡y más allá!”

Lluís-Esteve Casellas i Serra

Sección de Gestión Documental y Archivo y Administrador de Seguridad de Datos, Ayuntamiento de Girona
lecasellas@ajgirona.cat
.....

Resumen

El texto tiene una doble finalidad: exponer las pautas en la identificación, selección y preservación de datos auténticos con base en la experiencia del ayuntamiento de Girona, y reivindicar la necesidad de promover políticas de preservación de datos de manera cooperativa con las de seguridad de la información, proyectos *open data* y protección de datos de carácter personal.

También se destaca la poca atención que ha tenido la gestión de documentos para la preservación de datos y el convencimiento de que si los profesionales no intervenimos, difícilmente podremos tener un rol significativo en otros ámbitos con mayor visibilidad como los proyectos *open data* o *big data* de nuestras organizaciones.

Palabras clave: Preservación de datos, preservación de la autenticidad de datos, open data

Abstract

The article is twofold: to expose guidelines to identify, select and preserve authentic data based on the practical experience of the city council of Girona; and to assert the need to promote specific policies for data preservation. These policies should be made cooperatively with the domains of the security of information, open data projects and protection of personal data. Also, the article highlights the lacking attention that records management has been given so far to the data preservation. From the author viewpoint if professionals do not intervene decisively on data preservation, we can hardly play a significant role in other most visible areas such as open data or big data projects of our organizations.

Keywords: Data preservation, preservation of the data authenticity, open data

Introducción

En años recientes, la creación de repositorios de datos y conceptos como *big data*, *open data*, *data governance*, *linked open data*, *data management*, protección de datos o privacidad de datos, *data science*, *data mining*, son habituales en el contexto de cualquier organización pública o privada con cierto nivel de complejidad en su gestión diaria. En las últimas décadas, el incremento de la gestión de datos ha sido exponencial ya sea por la publicación y compartición de datos, la generación automática de datos mediante estaciones de captura sistemática (meteorológicas, tráfico, transporte público, consumo, etc.) o por el "internet de las cosas" –ya en desarrollo.

Desde un punto de vista de gestión administrativa, el incremento de datos ha sido espectacular, aunque probablemente con menor visibilidad. Implementar sistemas de gestión de bases de datos ligado a la automatización de los flujos de trabajo de las organizaciones ha contribuido a la creciente tendencia de "dejar de fijar" la información en documentos (información desestructurada) incluso en formato digital (por ejemplo, en PDF), para mantenerse en forma de datos (información estructurada) directamente en los sistemas que los gestionan. De este modo, progresivamente los sistemas se centran más en los datos que en los documentos, al menos tal como los conocíamos hasta ahora. Estamos, pues, ante la "deconstrucción" de los documentos y el aumento de la cuantía de los sistemas orientados a datos: los sistemas datacéntricos.

Sin embargo, más allá de sus componentes estratégicos para la organización una aproximación general a la gestión de datos, desde la perspectiva profesional de la archivística y –de forma concreta– de la gestión de documentos resulta bastante decepcionante. En apariencia, más allá del temor por la preservación de la memoria por parte de los archivistas clásicos, el sector profesional no ha tenido interés por la gestión y preservación de datos. Resulta aún más sorprendente cuando nos planteamos cuál debe ser nuestro rol en el nuevo contexto *big data* y *open data*, proyectos que ofrecen una mayor proyección y visibilidad a los actores que los gestionan pero con escasa participación

de los profesionales de los archivos y la gestión de documentos.¹ Nuestro rol, con frecuencia, menor en dichos proyectos contrasta con la necesidad de la perspectiva para su correcto desarrollo. Por ejemplo, cómo aplicar criterios de retención y de disposición acordes con los calendarios de conservación, pero también para asegurar la trazabilidad necesaria que garantice la autenticidad de los datos obtenidos y explotados;² sin duda, todo un reto profesional.

Por consiguiente la conclusión es que nuestra baja participación en los proyectos mencionados (voluntaria o motivada por terceros) es la consecuencia directa de la falta de atención que los profesionales hemos dado a la preservación de datos, entendida ésta como la preservación de cualquier otro tipo de documento digital, es decir, desde el inicio de su ciclo de vida y, por tanto, desde la fase de diseño de los sistemas que gestionan los datos.

Breve estado de la cuestión

Hasta ahora la mayor parte de iniciativas entorno a la gestión de datos estuvieron vinculadas con la explotación directa (*data management*) por parte de los profesionales de las tecnologías de la información, o bien, a preservar datos derivados de la investigación (*research data management*), principalmente bajo el liderazgo de bibliotecas universitarias o instituciones de investigación científica. Como resultado, la gestión de datos aparece ligada a nuevos perfiles profesionales como los

.....

¹ Page, Jessica, Flinn, Andrew, Shepherd, Elizabeth, "The Role of the Records Manager in an Open Government Environment in the UK", (FinalReport.), en InterPARES TRUST Project, noviembre, 2014, disponible en: <https://interparestrust.org/assets/public/dissemination/EU03_20141105_RoleRMOpenGovUK_FinalReport.pdf> [Consultado en febrero de 2016].

² Sobre los retos ante proyectos big data y open data: McDonald, John, Léveillé, Valerie, "Wither the Retention Schedule in the Era of Big Data and Open Data," Records Management Journal. Big data, open data, open source: information and records management opportunities and challenges, Anne Thurston, ed., vol. 24, núm. 2, 2014, pp. 99-121.

de *data manager*, *data scientist* o *data curator*, éste último vinculado al ámbito de la gestión de contenidos. Por tanto, se deduce que hasta el momento la preservación de datos se ha obviado o se ha reducido a la visión clásica y superada de intervenir una vez que los datos pasan al archivo o al centro responsable de su custodia. Por otra parte, el término *data* a menudo se utiliza en un sentido amplio y no de forma específica con referencia a datos estructurados. De igual manera puede ser de interés analizar las políticas y documentos de referencia del *Consortium of European Social Science Data Archive* (CESSDA),³ *Digital Curation Centre* (DCC),⁴ *UK Data Archive*,⁵ en el ámbito académico los de la *Canadian Association of Research Libraries* (CARL)⁶ o de la *University of Minnesota Libraries*,⁷ desde un punto de vista más tecnológico, las especificaciones técnicas de *Data Documentation Initiative* (DDI).⁸

El único ejemplo localizado con una perspectiva más alineada con la archivística corresponde a la política del *Canadian Geospatial Data Infrastructure*,⁹ orientada a los sistemas de información geográfica basada, entre otros, en el proyecto InterPARES 2 y el modelo OAIS. Sin embargo, dicha perspectiva corresponde a una visión más centrada en la preservación en fase inactiva y no en la captura de los datos y en el mantenimiento en fase activa.

.....

³ Disponible en: <<http://cessda.net/CESSDA-Training/Data-Archives-and-Digital-Preservation>> [Consultado en diciembre de 2015].

⁴ Disponible en: <<http://www.dcc.ac.uk/resources/how-guides>> [Consultado en diciembre de 2015].

⁵ Disponible en: <<http://www.data-archive.ac.uk/media/54776/ukda062-dps-preservationpolicy.pdf>> [Consultado en diciembre de 2015].

⁶ Disponible en: <http://carl-abrc.ca/uploads/pdfs/data_mgt_toolkit.pdf> [Consultado en diciembre de 2015].

⁷ Disponible en: <<https://www.lib.umn.edu/datamanagemen>>t [Consultado en diciembre de 2015].

⁸ Disponible en: <<http://www.ddialliance.org/>> [Consultado en diciembre de 2015].

⁹ GeoConnections, Hickling Arthurs Low Corporation, Geospatial Data Preservation Primer, disponible en: <<http://geoscan.nrcan.gc.ca/starweb/geoscan/servlet.starweb?path=geoscan/fulle.web&search1=R=296299>> [Consultado en diciembre de 2015].

En el ámbito de los archivos el proyecto *Electronic Resource Preservation and Access Network* (ERPANET, 2002-2004), de la Comisión Europea, analizó diversas experiencias en la preservación de bases de datos por parte de distintos archivos nacionales (Francia, Bélgica, Estados Unidos, Holanda, Inglaterra, Noruega y Suiza). A pesar de que algunos de los proyectos analizados iniciaron en la década de 1980, principalmente los basados en la simple extracción más o menos sistemática de datos en ficheros planos, se constataron como problemas a resolver el acceso a las copias de preservación, la evaluación de las bases de datos, la relación entre éstas y los documentos y la documentación de las propias bases de datos. Sin embargo, el punto más relevante desde el enfoque de este artículo es la constatación de que la preservación de las bases de datos se fundamentaba esencialmente en la perspectiva tecnológica, mientras que la archivística no había desarrollado una metodología propia.¹⁰ En este sentido, los *snapshots* o instantáneas extraídas de las bases de datos imposibilitaban detectar cambios en bases de datos dinámicas. También los vínculos entre bases de datos relacionadas que podían ser modificadas por separado enfrentaban problemas en cuanto a su preservación. De este modo, el punto clave se situaba en la preservación de los modelos relacionales de datos. En este punto, el proyecto suizo SIARD (*Software Invariant Archiving of Relational DataBases*) parecía ofrecer mayores ventajas pues permitía (y permite) procesos automatizados de extracción de datos de acuerdo con el modelo relacional de datos, manteniendo la información necesaria para recargar los datos en los sistemas de información. Además, la decisión para preservar una base de datos o una parte de ella se toma a partir del análisis de las funciones del productor con base en el valor evidencial e informativo de los datos, lo cual implica no tomar como

.....
¹⁰ ERPANET, The Long-term Preservation of Databases. ERPANET Workshop Report, Bern, abril 9-11, 2003, disponible en: <<http://www.erpanet.org/events/2003/bern/index.php>> [Consultado en diciembre de 2015].

base el momento de su transferencia al archivo sino una evaluación prospectiva de los sistemas de bases de datos.

A pesar del interés y relevancia de esta iniciativa cabe plantear algunas observaciones. La primera es de concepto sobre la necesidad o no de preservar los modelos relacionales de datos. Es evidente que en la decisión prevalece el componente tecnológico de mantener las funcionalidades de las bases de datos, no la singularidad de los documentos aún manteniéndolos en un formato que permitiera su explotabilidad. De esta premisa pueden derivar metodologías distintas en el análisis y selección de los conjuntos de datos, pues no queda claro que la prioridad sea la preservación del documento, entendido como su forma manifestada a través del sistema con todos sus atributos de identidad, integridad y autenticidad. Por ejemplo, cabría preguntar: ¿Lo que constituye a un documento es la tabla de datos o cada una de las vistas de la base de datos? ¿Cómo se gestiona la presunción de la autenticidad de dichos datos?¹¹

La segunda observación es en torno al acceso de los datos, puesto que su almacenaje en tres ficheros y formatos distintos (XML, SQL y texto plano UNICODE) requiere una interfaz de presentación, sea la de origen o no. Por consiguiente, surgen dudas desde el punto de vista de los usuarios potenciales sobre cómo acceder a la información. Finalmente, cabe cuestionar si preservar bases de datos relacionales tiene sentido en un contexto *open data* que implica una selección clara de los datos de manera que no dificulte su reutilización.

En conclusión, se puede afirmar que la preservación de las bases de datos constituye un severo déficit en el ámbito profesional del cual deriva una participación menor en los proyectos en auge, como son los contextos *big data* y *open data*.

.....
¹¹ Duranti, Luciana, Thibodeau, Kenneth, (Alejandro Delgado, trad.), "El concepto de documento en entornos interactivos, experienciales y dinámicos: la visión de InterPARES", originalmente en *Archival Science: International Journal on Recorded Information*, vol. 6, núm. 1, marzo, 2006, pp. 13-68, disponible en: <http://www.interpares.org/display_file.cfm?doc=ip2_dissemination_jar-s_duranti~thibodeau_archival_science_6_2006.pdf> [consultado en diciembre de 2015].

Hasta ahora la preservación de datos se ha centrado en los producidos en el ámbito de la investigación y con la aportación financiera principalmente del sector público. Las pocas experiencias e iniciativas sobre este ámbito desde el contexto de los archivos han sido lideradas por algunos archivos nacionales, aunque se constata un enfoque más tecnológico que metodológico. El objetivo de este artículo¹² es aportar una serie de reflexiones metodológicas y estratégicas en torno a la preservación de datos a partir de la experiencia del Servicio de Gestión Documental, Archivos y Publicaciones del Ayuntamiento de Girona¹³ en materia de preservación, de protección de datos de carácter personal y de datos abiertos.

¿Por qué una política específica sobre preservación de datos?

Antes de proseguir se debe considerar si es necesaria una política específica para la preservación de datos. El ayuntamiento de Girona dispone de un sistema de gestión de documentos razonablemente

.....
¹² Este artículo toma de referencia los siguientes trabajos del autor:

“Data Preservation, Open Data and Protection of Personal Data: a concurrent strategy in Girona City Council”, comunicación presentada en el 3a Conferencia Anual del Consejo Internacional de archivos: Archives: Evidence, Security and Civil Rights. Reykjavik, 2016.

“The mapping, selecting and opening of data: The records management contribution to the Open Data project in Girona City Council”, en Records Management Journal, vol. 24, ISS: 2, pp. 87-98, la traducción al castellano próximamente estará disponible en: <http://www.girona.cat/sgdap/cat/articles_sgdap.php>

Casellas, Lluís-Esteve, Oliveras, Sònia; Reixach, “The authenticity of Data-Centric Systems at the Girona City Council”, en InterPARES 3. (actualmente en proceso de revisión), disponible en: <http://www.interpares.org/display_file.cfm?doc=ip3_catalonia_cs02_final_report_EN.pdf> [Consultado en febrero de 2016].

¹³ La ciudad de Girona está situada en Cataluña, entre Barcelona y la frontera francesa, cuenta con 100 000 habitantes y su ayuntamiento con aproximadamente 1000 empleados.

eficiente desde hace varios años y también ha desarrollado una regulación específica para los documentos digitales, sin embargo, no se han llevado a cabo acciones orientadas a la preservación de documentos en bases de datos, ni siquiera la normativa sobre documentos digitales proporcionaba directrices para ello.

Conforme con lo comentado, los datos generados en la gestión diaria han aumentado debido a la automatización de los procesos de trabajo. Sin duda, en el marco de la administración electrónica la tendencia es la gestión basada en los datos como resultado del uso intensivo de los formularios en línea. Por lo tanto, la información puede fijarse en un documento en papel, digital (generalmente en PDF) o simplemente permanecer en el sistema –con el riesgo que esto representa–. Nos enfrentamos a una transición entre dos modelos de gestión de los documentos que incluso pueden combinarse: el modelo docucéntrico y el modelo datacéntrico, que definimos del siguiente modo:

- **Modelo docucéntrico.** Sistema de información en el cual la unidad de información mínima es un fichero (objeto digital), cada uno de los ficheros pueden tener un formato distinto. La comprensión y explotación individual de cada fichero no depende de un gestor de bases de datos, sino de la aplicación específica capaz de representar la información contenida en él.
- **Modelo datacéntrico.** Sistema de información estructurada en el cual la unidad de información mínima es el dato. Los datos tienen el mismo formato y dependen de un gestor de bases de datos para su comprensión y explotación.

Entender los datos dentro de un modelo datacéntrico se relaciona con la preservación del contexto de su creación y, sobre todo, con los formularios y las vistas o reportes que los complementan, por consiguiente surge un nuevo desafío: documentar el contexto de creación y la relación entre los sistemas de información.

Otros riesgos añadidos son:

- Cómo evitar la modificación de los datos si se actualizan constantemente.
- Cómo asegurar la preservación de datos sin ningún proceso definido para extraerlos de los sistemas de información.
- Cómo garantizar la presunción de autenticidad de los datos desde el momento en que se han registrado hasta su extracción del sistema.

Hasta hace poco la respuesta a estas cuestiones –desde un punto de vista tecnológico– era “no hay de qué preocuparse; transferiremos la base de datos completa”. Más allá de la posibilidad o la viabilidad de este tipo de transferencias, ¿realmente queremos transferencias de bases de datos completas? En el ayuntamiento de Girona la respuesta hasta el momento ha sido clara: “No, gracias”.

Por lo general, los sistemas de información generan o contienen documentos. En el contexto actual considerar una base de datos como un documento individual debería ser un hecho excepcional. El presente artículo se centra en los documentos contenidos en los sistemas de bases de datos, los cuales se extraen bajo la forma documental de *registro*, similarmente a sus precedentes en papel: los *libros de registro*. El desafío es, por lo tanto, la identificación de qué conjuntos de datos constituyen documentos o puede ser considerados como tales, qué requisitos deben satisfacer y cuáles estrategias deben adoptarse para preservarlos adecuadamente.

Desde la archivística resulta indispensable tener en cuenta las siguientes premisas como aspectos esenciales:

1. Fijar la información, sobre todo la contenida en sistemas permanentemente actualizables y actualizados (*registros continuos*) y, en particular, aquella evaluada de conservación permanente.
2. Garantizar la autenticidad de los datos en dos etapas: la primera, desde el momento de captura de la información hasta su disposición; la segunda, a partir de su disposición en adelante (cuando implique su extracción del sistema). Resulta

evidente que la principal dificultad se encuentra en la primera de las dos etapas.

3. Mantener la explotabilidad de los datos ya que no tendría sentido preservarlos sin la opción de explotarlos, principalmente desde el punto de vista de la organización.

Se define explotabilidad de datos como la capacidad de un conjunto de datos extraídos de un sistema de información para preservar su valor informativo y aprovechar su potencial fuera del sistema en que fue creado. Esta capacidad se vincula con la descripción y la documentación de la estructura de datos, del contexto de producción y del contenido, así como con el mantenimiento de los requisitos de identidad, integridad, autenticidad y usabilidad de los datos.

La identificación de los registros, la selección de datos

La identificación de los registros, entendidos como documentos que recapitulan la información, es la primera clave para la preservación de gestión de la información en las bases de datos. Así, los productos procedentes de bases de datos pueden ser de dos tipos:

- **Documentos generados**, basados en la combinación de datos y plantillas de texto fijados mediante su impresión en papel o en un fichero digital, generalmente PDF.
- **Documentos contenidos** en grupos de datos vinculados y coherentes, creados y mantenidos en las bases de datos. Estos grupos de datos acumulados secuencialmente constituyen o pueden constituir registros. En este caso, es indispensable identificar la función y las actividades que los generan y la descripción del contexto en el que se crearon.

Por lo tanto, el primero corresponde a un modelo docucéntrico y el segundo, a uno datacéntrico.

Las características generales de los documentos generados en un modelo docucéntrico son:

- La unidad mínima de información es un fichero o un objeto digital, no necesariamente en el mismo formato.
- Los documentos se fijan de forma individual, ya sea como un solo objeto o vinculados con más de uno.
- La autenticidad se basa principalmente en algún tipo de autenticación basado en firma digital.
- La gestión de los objetos digitales está vinculada pero no integrada en la base de datos.
- La facultad de aprovechar de forma estructurada de estos documentos no es una función requerida, aunque los datos utilizados para crearlos puedan pertenecer a un registro formalmente definido.

Las características generales de los documentos contenidos en un modelo datacéntrico son:

- La unidad de información mínima es el dato, siempre en el mismo formato.
- Los datos no se extraen necesariamente del sistema.
- Los datos se crean y se acumulan de forma secuencial y pueden constituir registros.
- Su comprensión es únicamente posible a partir de las relaciones entre los datos y las plantillas o vistas previamente definidas.
- Los registros gestionados en las bases de datos pueden ser definidos e identificados formalmente y declarados como documentos.

A partir de estas consideraciones se distinguen dos tipos de registros de acuerdo con la naturaleza de sus asientos o entrada de datos:¹⁴

- **Registros que documentan acciones:** el objetivo de una acción es validarla mediante una entrada en el registro, que puede implicar la asunción de derechos u obligaciones, según el caso, por ejemplo, en el padrón municipal de habitantes.
- **Registros que controlan otros documentos:** cada entrada hace referencia a un documento pero su asiento en el *registro* no sustituye su valor o el de la acción que documenta, por ejemplo, de un registro de facturas.

Por otra parte, la relación entre asientos de entrada determina la delimitación temporal del registro como documento. En este sentido, los registros pueden ser continuos o discontinuos en el tiempo:

- **Registro continuo:** documento¹⁵ formado por la acumulación secuencial de asientos que por su naturaleza constituyen cada uno de ellos un documento, de manera que en conjunto son una unidad documental compuesta. Un registro continuo se caracteriza por no estar formalmente delimitado en el tiempo y por el hecho de que cada asiento es la culminación de la tramitación que lo motiva, por tanto, la validación individual del asiento produce los efectos derivados de la función de registro, ya sean inmediatos o condicionados, a un periodo de carencia, conforme a su regulación legal, es decir, el asiento contiene el valor legal de la acción y asume el valor de los documentos aportados como prueba de soporte; por ejemplo, el padrón

.....
¹⁴ Se empleará "entrada" o "asiento" para evitar repetir "registro" y causar confusión.

¹⁵ Antes de las bases de datos estos registros podían presentarse por acumulación de libros de registro (de la propiedad, mercantiles, de hipotecas, etc.) o de expedientes de inscripción, el conjunto de los cuales constituía el registro.

municipal de habitantes, el registro de actividades económicas, espectáculos y salud pública, el registro de infraestructuras críticas, o el registro curricular de los alumnos.

- **Registro discontinuo:** documento¹⁶ formado por la acumulación secuencial de asientos en un periodo predefinido y concreto; su conjunto constituye una unidad documental simple. Generalmente, un asiento de un registro discontinuo no asume el valor de los documentos que lo justifican sino que lo cobra en razón de ellos para su control administrativo y la formalización final. Por consiguiente, en los registros discontinuos el valor legal reside en el conjunto de los asientos y en los documentos que los soportan; por ejemplo, los registros contables, los padrones fiscales o el registro de matriculación de alumnos.

La comparación entre las características de uno y otro tipo ayuda a comprender la importancia de documentar el contexto en el que se crean los registros e incluso su identificación. Si los registros no son formalmente determinados es probable que su función implícita no se identifique, en especial en los continuos, por tanto, su preservación puede verse comprometida si tiene que ser permanente o por un periodo largo.

Existe otro factor de riesgo a tener en cuenta: estos registros suelen compartir datos con otros, por lo cual la identificación no puede limitarse simplemente a la función, la actividad y los datos que apoyan el documento; se debe identificar y describir la relación entre los diversos sistemas de bases de datos para garantizar su preservación en todos y cada uno de sus contextos de creación.

.....

¹⁶ Antes de las bases de datos estos registros se presentaban tradicionalmente en libros de registro anuales.

En este sentido, se podría inferir que los registros que controlan otros documentos tienen un interés secundario, sin embargo, constituyen una herramienta indispensable para documentar el contexto y la trazabilidad implícita de los procesos de gestión. La identificación, el conocimiento y la descripción de las relaciones y los vínculos existentes entre los documentos son esenciales para la preservación de documentos digitales y, en consecuencia, este tipo de registros contribuye a garantizar la confianza en el sistema.

En cuanto a la selección de los datos que tienen que conformar el registro, si previamente existía un precedente en papel en formato de libro o similar, la selección se realiza con base en los datos que conforman el documento original, por ejemplo el padrón municipal de habitantes. Si la función o la actividad gestionada en las bases de datos se refleja (o se reflejaba) en un expediente, la selección se basa en la identificación de los principales datos del proceso completo de tramitación y, al mismo tiempo, en la de los datos cualitativos que diferencian esta tramitación de otra de la misma naturaleza, lo que se obtiene es un registro más completo que el tradicional catálogo de expedientes, pues se añaden datos adicionales con valor informativo. Un ejemplo aplicable sería el de las obras de particulares sujetas al régimen de comunicación previa. La experiencia en ambos procesos de selección nos ha llevado a consensar con el departamento de informática como principio de actuación el potenciamiento de la creación y formalización de registros administrativos específicos de los procesos administrativos, como un desarrollo complementario del registro general de expedientes. Esta metodología nos permite plantear tres tipos de intervención:

- **Acciones correctivas** en relación con registros almacenados en bases de datos con anterioridad, incluso de bastantes años atrás, por ejemplo, el padrón municipal de habitantes, un registro continuo computarizado a partir de 1996.
- **Acciones preventivas**, por la incorporación de estas pautas de actuación en las rutinas de gestión de datos en los sistemas actuales. En algunos casos, esto ha permitido crear de cero

registros recapitulativos de otros documentos en papel, lo que ha facilitado su sustitución y eliminación regulada.

- **Acciones predictivas**, es decir, la evaluación previa y la identificación de los datos de los procesos de gestión susceptibles de conformar registros a partir del análisis de las funciones y actividades y, si fuera necesario, su regulación institucional, ante cualquier propuesta de diseño o rediseño de los sistemas.

La localización de los conjuntos de datos

Recopilar los datos susceptibles de ser evaluados puede ser otro inconveniente. En este sentido, una primera identificación global de las funciones, actividades y productores es posible mediante:

- El **cuadro de clasificación de documentos**: proporciona información de documentos creados por el ayuntamiento pero también de las unidades administrativas que los crean y cómo lo hacen. Permite identificar series documentales que corresponden a registros administrativos, formalizados o no.
- El **registro general de expedientes**: ofrece datos básicos a nivel de catálogo de expedientes, sin embargo, en series documentales que han implementado procesos de automatización de flujos de trabajo es posible obtener datos cualitativos complementarios que pueden resultar de interés. Desafortunadamente, estas series representan una minoría en relación con el conjunto del cuadro de clasificación.

Estos instrumentos proporcionan una perspectiva global de la organización difícil de lograr por otros medios, aunque es importante remarcar que presentan limitaciones y que se requiere información más detallada. Dos instrumentos complementarios pueden ser el catálogo de aplicaciones informáticas y el registro de ficheros con datos de carácter personal, dado que la privacidad de datos es también una de las funciones de nuestro servicio. Desde la perspectiva de la gestión de documentos, el catálogo de aplicaciones informáticas se orienta

fundamentalmente a la vinculación de los desarrollos informáticos específicos con las funciones y las actividades de la organización y, también, a la identificación del documento o grupo de documentos que producen. Cabe remarcar la diferencia entre el punto de vista archivístico sobre el catálogo de la gestión de documentos y el punto de vista tecnológico, más orientado en el tipo de procesos y la gestión de la propia herramienta tecnológica.

Por consiguiente, el objetivo principal del catálogo de aplicaciones es identificar los datos en los sistemas de información que constituyen documentos, con el fin de facilitar su gestión y preservación. El catálogo fue elaborado a partir del inventario de aplicaciones del departamento de informática y se complementó con entrevistas con el personal técnico responsable de cada unidad de producción y, en ocasiones, con el personal administrativo; los resultados están alineados con el cuadro de clasificación de documentos.

El registro de ficheros con datos personales es de uso obligatorio y auditable para todos los ficheros susceptibles de ser declarados e inscritos en la Agencia Catalana de Protección de Datos. El registro ofrece información complementaria, porque detalla la estructura de los datos personales dentro de cada fichero; se elaboró a partir de encuestas y entrevistas con los interlocutores designados por cada unidad administrativa con el fin de cumplir con la legislación sobre protección de datos. El resultado final se alineó con el cuadro de clasificación de documentos, aunque inicialmente no con el catálogo de aplicaciones. El mapeo de datos exhaustivo en toda la organización municipal constituye actualmente para el departamento de gestión documental una de las líneas de estrategias de desarrollo del proyecto.

La gestión de la presunción de autenticidad

Las propuestas presentadas parten de un caso de estudio del Proyecto InterPARES 3, liderado por la Universidad de la Columbia Británica (Vancouver, Canadá) y centrado en la preservación de la autenticidad en los sistemas datacéntricos del Ayuntamiento de Girona, principalmente sobre su aplicación concreta al padrón municipal de

habitantes, un registro continuo evaluado de conservación permanente calificado como documento esencial de la organización.

Al hablar de registros –sobre todo si sus datos nunca han sido extraídos de los sistemas–, el tratamiento de la autenticidad implica la gestión continua de las medidas adoptadas para preservar los datos en línea, sin embargo, si extraemos los datos del sistema periódicamente, el tratamiento de la autenticidad tiene dos etapas: la primera, desde el momento de captura de la información hasta su disposición; y la segunda, a partir de su disposición en adelante, cuando implique su extracción del sistema y no su eliminación. La extracción de los datos no impide el mantenimiento de la autenticidad de los datos en línea, pero la posibilidad de poderlos comparar refuerza –por redundancia– la autenticidad del sistema debido al efecto espejo, por lo tanto, el sistema se vuelve más fiable.

En realidad, la gestión de la autenticidad de los datos extraídos es más fácil que la de los datos en línea, debido a que para validarlos podemos utilizar sistemas de autenticación como la firma digital. Para los datos en línea se consideraron distintos sistemas de validación para garantizar de forma individualizada la información de los asientos de entrada: la firma digital, la aplicación de una función de HASH, e incluso la aplicación de una función HASH encadenada si el nivel de criticidad de los datos lo requería. Sin embargo, estas propuestas fueron descartadas porque conllevaban una gestión más compleja de los asientos y del registro, en general, sin mejoras significativas. También se planteó la posibilidad de diseñar un sistema independiente para la duplicación de los datos seleccionados, gestionado por un tercero de confianza aunque a nivel interno, el departamento de gestión documental. No obstante, la extracción sincronizada de datos con controles adicionales para el aseguramiento de su integridad complicaba más la operación en lugar de mejorarla.

Finalmente, asumimos que una gestión cuidadosa de los usuarios a nivel de privilegios de acceso y roles de mantenimiento garantizaría el uso y manipulación pertinente de la información, además de las correspondientes e indispensables medidas de seguridad informática de soporte; aun así se consideró como una condición relevante el

mantenimiento de los *logs* de creación y validación de los asientos de entrada durante todo el periodo de retención del documento, mientras que los *logs* limitados al acceso de consulta podrían eliminarse en el corto plazo. En cualquier caso, la presunción de confianza en el sistema y en las personas que lo manejan parecía inevitable al mismo tiempo que la implementación de medidas organizativas y de control tecnológico que garantizaran la identidad, la integridad y la autenticidad de los datos.

La autenticidad de los datos extraídos del sistema se basa en la política de preservación, en la aprobación de un procedimiento específico para su extracción, su incorporación al Sistema de Gestión de Registros y su autenticación mediante firma digital. En términos de cumplimiento legal, el ayuntamiento aprueba procedimientos específicos para formalizar y obtener de cada registro en relación con:

- La identificación del conjunto de datos seleccionado que constituye el registro y que se declara formalmente como documento.
- La descripción de la función y las actividades que el registro representa, vinculadas con el cuadro de clasificación de documentos.
- El proceso que se aplicará para extraer los datos: qué se pretende hacer; quién puede hacerlo; quién puede validarlo; cuándo y cómo se va a hacer (QQCC: qué quién, cuándo, cómo).
- Los formatos de fichero de los datos extraídos.
- La incorporación del registro al sistema de gestión de documentos y su autenticación mediante una firma digital.

En cuanto a los formatos de preservación, los conjuntos de datos se extraen en PDF y TXT. La elección de la opción PDF se basa en facilitar la comprensión primaria del registro, mientras que el TXT está destinado a mantener la explotabilidad de los datos en el tiempo. La descripción de la estructura de datos, de las especificidades de cada campo, su contexto y sus vínculos son absolutamente indispensables.

Cuadro 1 Recomendaciones generales aplicadas para la preservación de la presunción de la autenticidad de los datos

Integridad

- Identificación y definición del registro.
- Robusta gestión general de los usuarios.
- Actualización por adición de datos, no por sustitución.
- Fijación periódica y regulada por extracción sistemática de datos.
- Integración en el Sistema de Gestión de Documentos.
- Firma digital sobre los datos extraídos.
- Verificación de firma digital por la Autoridad de Certificación.

Confiabilidad

- Regulación precisa de la gestión del registro.
- Definición de las responsabilidades de gestión.
- Preservación de *logs* de creación acorde con los periodos de retención definidos.
- Regulación institucional de los procedimientos.

Autenticidad

- Cumplimiento de los criterios de integridad y fiabilidad.
- Aseguramiento de la no alteración de los datos en el largo plazo.
- Identificación de otros documentos relacionados con la base de datos: trazabilidad implícita.
- Extracción periódica, sistemática y regulada de datos.
- Asignación de responsabilidades para la extracción de datos al Departamento de Gestión Documental y Archivo.
- Continuidad en el tiempo de estas funciones y repetibilidad de las acciones con variabilidad acotada de los resultados.

Usabilidad

- Formatos de preservación apropiados.
- Garantía sobre la explotabilidad de los datos.

Desde el punto de vista de seguridad los registros calificados como esenciales son también fijados en microfilm, a partir de la impresión directa del PDF debido a la naturaleza duradera del medio y la calidad óptima de impresión.

En conclusión, la preservación de datos del Ayuntamiento de Girona se basa en dos premisas: la *securización* de los datos en línea; y, la *duplicación* de datos mediante la extracción periódica de conjuntos de datos bajo la forma documental de registros, previamente identificados, definidos y declarados como registros administrativos (Cuadro 1).

Preservación de datos, seguridad de la información, *open data* y protección de datos de carácter personal: ¿una estrategia concurrente?

En un contexto en el que los datos parecen tener un papel determinante ¿puede plantearse una política para preservar datos de forma aislada de otras políticas? Quizá no sería razonable por motivos de eficiencia o por la diferencia en el peso específico de cada ámbito apuntado en el título del apartado. Por una parte la relación entre *open data* y la protección de datos personales es bastante obvia, ya que la disociación o anonimización de los datos sensibles es prácticamente inevitable. Por otro lado, la relación entre preservación y seguridad también parece bastante clara, pero principalmente porque la primera requiere de la segunda. El problema para la preservación de datos es que no es percibida como una necesidad para la organización, mientras que la seguridad y la protección son inevitables, bien sea por obligación legal o para evitar pérdidas o daños a la organización. Tanto la seguridad como la protección de datos –y también las iniciativas *open data*–, tienen una mayor visibilidad en el seno de las organizaciones que no tiene la preservación de datos, entendida la mayor parte de las veces como mera acción de preservación de la memoria.

No obstante, desde la experiencia en el Ayuntamiento de Girona consideramos que la visión global de la gestión de documentos pone de relieve la necesidad de vincular las políticas en estos ámbitos con un doble objetivo:

- Ser más eficaces y eficientes en cada uno de estos ámbitos mediante políticas cooperativas.
- Promover el rol de los archiveros en la gestión de datos de la organización, principalmente en proyectos *open data*, dado que son más visibles para la alta dirección.

Sin embargo, reforzar el rol del archivero en la gestión de datos no es sólo una cuestión de visibilidad, se trata de una estrategia para fortalecer políticas y procedimientos específicos que tienen una influencia significativa en la preservación de datos desde una perspectiva archivística, por lo que es indispensable identificar quién hace qué en el seno de la organización e identificar los puntos de coincidencia (Figura 1).

Cada ámbito de trabajo tiene instrumentos específicos para su gestión –como se comentó–, se tiene de forma obligatoria un registro de ficheros con datos de carácter personal vinculado con un nivel más operativo con el inventario de activos de información, también preceptivo y ligado a la seguridad informática de la organización. Cualquier proyecto *open data* requiere de un catálogo de datos abiertos con los conjuntos publicados. Por último, el instrumento indispensable desde la perspectiva archivística es el catálogo de aplicaciones informáticas, que vincula los activos a las funciones y actividades descritas en el cuadro de clasificación de documentos.

Cada uno de estos ámbitos se administra por separado, aunque de una manera u otra en realidad se relacionan (Figura 2). Sin embargo, podrían ser gestionados de una manera más coordinada porque comparten algunas necesidades importantes:

1. La identificación y el mapeo de los activos de información de la organización. Por activos se entiende a los conjuntos de datos gestionados en los sistemas de información ("servicios" desde el punto de vista de la seguridad).
2. Un conocimiento más detallado y la descripción de la estructura de datos de todos los activos de información.
3. La identificación y selección de datos de valor que son susceptibles de ser extraídos de los sistemas de gestión para su reutilización como datos abiertos y su preservación como registros.
4. La aplicación de la Evaluación de Impacto de Privacidad (PIA) para definir procedimientos de disociación sin pérdidas de valor informativo.
5. La aplicación de requisitos que garanticen la fiabilidad, exactitud y autenticidad de datos en línea y, también, de los extraídos del sistema.

En nuestra opinión, la perspectiva archivística puede dar una respuesta eficaz a estas necesidades compartidas, por lo que la propuesta de cooperación se basa en el análisis de los activos de información desde el punto de vista de la gestión del ciclo de vida con el objetivo de establecer una política de preservación de datos y, al mismo tiempo, también resolver otras necesidades del ayuntamiento. Las acciones derivadas de la gestión del ciclo de vida de los activos de información pueden determinarse mediante:

- **Identificar los activos de información** a través del catálogo de aplicaciones, el cual a su vez incorporaría la descripción de sus estructuras de datos.
- **Evaluar los activos de información** mediante el análisis de las funciones y actividades que motivan la creación de los conjuntos de datos.
- **Preservar la autenticidad** mediante la aplicación de procedimientos y requisitos formales y técnicos que documenten

Figura 1 ¿Quién hace qué?: puntos de coincidencia

Departamento (*grado de influencia)	Instrumento	
Transparencia ***	Catálogo de datos abiertos	
Gestión de documentos*	Catálogo de sistemas de negocios ↓	Cuadro de clasificación →
Gestión de la privacidad (Depto GD)**	Catálogo de activos de información (GD) ↓ Registro de activos declarados con datos personales ↑	
Tecnologías de la información ****	Registro de activos de información (TI-Seguridad) → Inventario de software →	Registro de activos de información (GD) Catálogo de sistemas de negocios

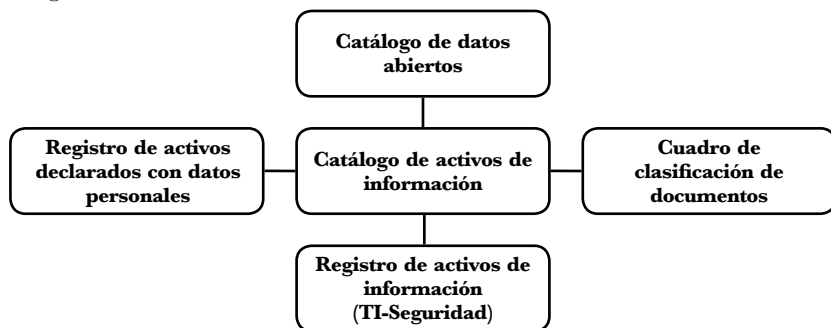
y garanticen el contexto, la identidad y la integridad de los conjuntos de datos definidos como documentos.

- Por último, derivado de la preservación de la autenticidad, **extraer datos** de forma regular y sistemática bajo la forma documental de **registro**, y su incorporación al sistema de gestión de documentos de la organización.

Si bien en términos de seguridad informática la incorporación de dichas acciones de gestión del ciclo de vida contribuye a identificar, dimensionar y mitigar riesgos, ¿cuál sería el valor añadido para los proyectos *open data* y la gestión de la protección de datos? Apuntamos algunos beneficios:

- Un control más efectivo y completo de los activos que contienen datos de carácter personal, el mapeo de activos reutilizables en un contexto *open data* y la identificación de los datos sensibles que deberían disociarse.
- La evaluación permite la selección de los datos de valor que constituyen un registro administrativo o que pueden constituirlo mediante una formalización previa, así como sus características formales y técnicas.
- La selección de datos facilita una identificación precisa de los

Figura 2 ¿Quién hace qué?: el escenario ideal u otra forma distinta de gestionar



datos relevantes que podrían publicarse así como la evaluación del riesgo asociado a su divulgación.

- La aplicación de los calendarios de retención sobre estos activos y de los procedimientos para su correcta disposición refuerzan la conformidad legal en relación con la protección de datos.
- La fiabilidad de los conjuntos de datos abiertos resulta reforzada al implementar los requisitos para preservar la autenticidad de las fuentes.
- Los procesos de extracción regulados y normalizados también se pueden aplicar para obtener productos de datos abiertos, reforzando así su confiabilidad.
- Los procedimientos de disociación de datos personales podrían ser incluidos en la fase de diseño de los procesos de extracción, de acuerdo con los criterios de privacidad por diseño.

Por lo tanto, la implementación de criterios archivísticos para gestionar el ciclo de vida de los activos de información presenta varios beneficios para la organización, concretamente, la evaluación archivística y la preservación de la autenticidad en los sistemas de bases de datos tienen un alto valor añadido con respecto a otras políticas que afectan a la gestión de datos en general.

Conclusiones

Hoy en día, somos testigos de un proceso de deconstrucción de los documentos, con el evidente riesgo para los datos que los constituyen, de acabar diluidos en los sistemas de información. Por esta razón, en el marco de una política general de gestión de documentos, una política específica de preservación de datos debe proporcionar a los documentos la resiliencia suficiente para que, a pesar de su presentación como datos, no pierdan su condición de documento.

En un contexto acelerado de *datification*, la apertura de datos, la protección de la privacidad y la seguridad de la información adquieren una relevancia y valor significativos –incluyendo el económico–. En contraste, la preservación de datos no se percibe de la misma

forma, por lo que debemos participar activamente en su preservación e incorporarla a políticas más consolidadas; el valor añadido de la evaluación y la preservación de la autenticidad puede contribuir a una mayor cooperación entre dichas políticas. Si no nos involucramos, difícilmente tendremos un rol relevante en los proyectos *open data* o *big data*.

Por lo tanto, más allá de algunas experiencias puntuales y dispersas como la que se ha expuesto, sería necesario promover iniciativas internacionales desde una perspectiva archivística con el fin de lograr modelos y metodologías de referencia en el campo de la preservación de datos para que los profesionales no tuvieran que "reinventar constantemente la rueda" (Anexo). Por consiguiente, abróchense los cinturones y a la preservación de datos... ¡y más allá!

Anexo. *Open data* y gestión de documentos en el Ayuntamiento de Girona ¿qué hemos aprendido?

1. EL CUADRO DE CLASIFICACIÓN DE DOCUMENTOS PROPORCIONA INFORMACIÓN VALIOSA PARA DETECTAR CONJUNTOS DE DATOS SUSCEPTIBLES DE SER PUBLICADOS, SOBRE TODO SI SE IDENTIFICAN LAS APLICACIONES ESPECÍFICAS QUE PRODUCEN LOS DOCUMENTOS.

En este sentido, el cuadro de clasificación de documentos no puede limitarse a su concepto tradicional, sino que tiene que ser el núcleo de un sistema de clasificación más complejo. La clasificación debe ir más allá del nivel de la serie documental para identificar el tipo de documentos que constituyen cada proceso de la gestión administrativa, por lo tanto, el cuadro de clasificación de documentos debe estar integrado en los sistemas de negocio, sólo así será posible identificar qué datos están asociados con la producción de cada documento. Por otra parte, este nivel de detalle facilita la identificación de los conjuntos de datos que constituirán directamente un documento. El cuadro de clasificación de documentos, por consiguiente, debería ser el vínculo entre la función, el documento y los datos relacionados para la creación del documento.

2. EL CATÁLOGO DE APLICACIONES ES MUY ÚTIL SI ESTÁ VINCULADO A LAS FUNCIONES Y LAS ACTIVIDADES DEL CUADRO DE CLASIFICACIÓN. EN ESTE CASO, SE REQUIERE LA INCLUSIÓN DE LA ESTRUCTURA DE LOS DATOS ESENCIALES SELECCIONADOS Y SU DESCRIPCIÓN.

El control sobre las aplicaciones que crean documentos es muy importante, sin embargo es esencial establecer un control a un nivel más detallado por lo que no es suficiente saber qué documento produce cada aplicación, también es necesario identificar qué datos son importantes para la producción de documentos. La identificación de estos datos facilitaría al mismo tiempo la identificación de los productos potenciales para proyectos open data. Además, en algunos casos estos datos se habrán constituido o podrán constituir al mismo documento, en cualquier caso, es crucial describir los datos desde un punto de vista tecnológico. En este sentido el Catálogo de Aplicaciones debe estar estrechamente vinculado a la gestión del cuadro de clasificación de documentos ya que son complementarios. Si uno proporciona información acerca de la función y los documentos producidos, el otro debe proporcionar información sobre si se producen en sistemas docucentricos o en sistemas datacentricos.

3. LA DEFINICIÓN DEL PERIODO CRONOLÓGICO DE LOS CONJUNTOS DE DATOS A PUBLICAR Y EL MANTENIMIENTO DE LOS FICHEROS DE DATOS ABIERTOS PUEDEN REFORZAR EL ROL DE LOS ARCHIVOS HISTÓRICOS TRADICIONALES.

Los productos publicados de datos abiertos son fáciles de usar y manipular por parte de los usuarios finales y no presentan ningún problema con respecto de las restricciones de privacidad de datos. Por lo tanto, ya que uno de sus objetivos es que sean accesibles permanentemente, es necesario tener en cuenta dos cuestiones: la necesidad de prever cómo clasificar y organizar estos productos de una manera coherente y eficaz; en segundo lugar, considerar la posibilidad de que, en cierto modo, este tipo de producto podría ocultar las fuentes originales custodiadas en los archivos históricos.

El aumento de productos de datos abiertos creados año tras año podría suponer una dificultad en gestión de la transparencia de las organizaciones, en consecuencia, parece lógico distinguir entre nuevos y antiguos conjuntos de datos y sería más eficaz gestionar y preservar los conjuntos de datos no actualizados en los archivos

históricos y publicarlos en sitios web lo que permitiría obtener una mayor visibilidad en el nuevo contexto de la gestión de datos.

4. LAS ESTRATEGIAS DE PRESERVACIÓN DE DOCUMENTOS EN BASES DE DATOS HACEN MÁS FÁCIL LA SELECCIÓN DE DATOS PARA PROYECTOS OPEN DATA, ASIMISMO, PUEDEN SER UN INCENTIVO IMPORTANTE PARA DESARROLLAR O REFORZAR ESTAS ESTRATEGIAS.

Los procesos de extracción de datos son esenciales en un proyecto open data. Por lo tanto, si una organización dispone cualquier estrategia de preservación de documentos en bases de datos resulta evidente que las selecciones anteriores hechas para este propósito pueden utilizarse directamente para desarrollar productos de datos abiertos. No obstante, cabe la posibilidad de una lectura en sentido inverso, es decir, los procesos de extracción de datos para proyectos open data podrían ser utilizados como una medida de preservación.

Cabe señalar que ciertos procesos de extracción pueden ser complejos y, a menudo, la organización puede considerar que el desarrollo de estrategias para la preservación de datos no es un objetivo fundamental. Por el contrario, la publicación de datos abiertos es una clara tendencia, por lo que es más fácil asignar recursos humanos y económicos para dichos proyectos, por lo que deben considerarse como una oportunidad para incluir la perspectiva de la preservación, de hecho, su combinación es mucho más eficiente desde la perspectiva de la gestión de recursos debido a que el mismo diseño puede aportar beneficios en dos ámbitos simultáneamente. Por otra parte, la gestión de documentos puede contribuir significativamente a la selección de los datos más relevantes a partir de su experiencia acumulada en la evaluación de documentos.

5. LA INCLUSIÓN DE LA PERSPECTIVA OPEN DATA EN EL ANÁLISIS PREVIO A LA AUTOMATIZACIÓN DE FLUJOS DE TRABAJO PODRÍA FACILITAR LA EXTRACCIÓN DE DATOS EN EL FUTURO, DE HECHO, PERMITIRÍA IMPORTANTES AHORROS ECONÓMICOS Y DE RECURSOS Y EVITARÍA LA REALIZACIÓN DE ANÁLISIS ESPECÍFICOS POSTERIORES.

Como se mencionó ciertos procesos de extracción pueden ser complejos, pero puede ser resultado de tener que aplicar soluciones a aplicaciones ya en funcionamiento. Por esta razón, puede ser más fácil e incluso más barato predecir los procesos de extracción de datos en la fase de diseño de las aplicaciones. Conforme con lo

referido en el punto anterior, la experiencia en la evaluación de los documentos será relevante en la selección de los datos más significativos también en esta fase.

6. LA EVALUACIÓN DE LOS DATOS SIGNIFICATIVOS DE LA TRAMITACIÓN DE UN EXPEDIENTE PODRÍA SER ÚTIL PARA POTENCIAR LA CREACIÓN Y FORMALIZACIÓN DE REGISTROS ADMINISTRATIVOS QUE PODRÍAN INTEGRARSE EN LAS ESTRATEGIAS DE PRESERVACIÓN DE DATOS.

Como se mencionó la evaluación aplicada a los datos de una tramitación administrativa es directamente útil tanto para la preservación como para la publicación de conjuntos de datos abiertos. Sin embargo, la evaluación de los datos puede tener otra aplicación: creación de nuevos documentos como resúmenes o recapitulativos de otros documentos. Independientemente de si los sistemas se centran en los documentos o en los datos, los procesos de gestión informatizados generan datos; a menudo éstos constituyen el núcleo de la información de los documentos producidos en los sistemas de gestión, por lo que la selección y la documentación de su contexto de forma adecuada puede permitir a la organización la formalización de registros administrativos a partir de los datos acumulados en las tramitaciones gestionadas en los sistemas de información. En estos casos, la formalización y la preservación de estos registros pueden facilitar la destrucción de otros documentos una vez que su periodo de retención haya terminado. Al mismo tiempo, la publicación de conjuntos de datos relacionados con registros formalizados refuerza la autenticidad de los datos publicados desde la perspectiva del usuario final.

7. LA CALIDAD DE LOS DATOS ES UNO DE LOS PROBLEMAS MÁS RELEVANTES DE LOS PROYECTOS DE DATOS ABIERTOS Y SÓLO ES POSIBLE RESOLVERLO EN EL MOMENTO DE CREACIÓN Y CAPTURA DE LA INFORMACIÓN. LA GESTIÓN A POSTERIORI DE LA CALIDAD DE LOS DATOS PUEDE CONVERTIRSE EN UN PROCESO COMPLEJO Y EN CIERTO MODO "TRAICIONA" LA FILOSOFÍA *open data* DE PUBLICAR DATOS EN BRUTO O CON LA MÍNIMA MANIPULACIÓN POSIBLE. DE HECHO, LA ÚNICA OPCIÓN VIABLE ES VINCULAR LA CALIDAD DE DATOS CON LOS SISTEMAS DE GESTIÓN DE DOCUMENTOS

Los sistemas de gestión de documentos también tienen problemas de calidad en sus datos, por lo que relacionar los sistemas con los proyectos de datos abiertos no

resuelve estos problemas. Sin embargo, los conjuntos de datos abiertos a publicar generalmente o provienen directamente de las sistemas de gestión de documentos o de sus mismas fuentes, en consecuencia, establecer estrategias comunes parece lo más lógico. Los sistemas de gestión de documentos son un primer filtro de control de calidad por la misma inmediatez en la creación y uso de documentos que suponen, la creación de documentos apropiados y de calidad es un objetivo primario y si esto está garantizado será mucho más fácil extraer conjuntos de datos de calidad. En cualquier caso estas medidas siempre tendrán que producirse al inicio del proceso, en el momento de captura de la información. Cabe añadir que, además de la formación del personal hay otras medidas para mejorar la calidad tales como el diseño de sistemas de entrada de datos automática o semiautomática o incluso la posibilidad de que los usuarios visualicen en pantalla qué campos van a ser públicos y cuáles de ellos requieren especial protección por cuestiones de privacidad.

8. EN ALGUNOS CASOS LA DISOCIACIÓN DE DATOS PUEDE SER DIFÍCIL. EN ESTE SENTIDO, LA GESTIÓN DE DOCUMENTOS PUEDE CONTRIBUIR MEDIANTE ANÁLISIS ESPECÍFICOS Y LA INTRODUCCIÓN DE DIRECTRICES Y PAUTAS SOBRE LA ENTRADA DE INFORMACIÓN, FACILITANDO ASÍ LA DISOCIACIÓN AUTOMÁTICA POSTERIOR.

La mayoría de los proyectos open data describen en detalle los conjuntos de datos publicados desde el punto de vista de la estructura de datos, pero la información sobre el contenido suele ser muy limitada. Por otra parte, la fuente original ocasionalmente es descrita más allá de la inclusión del título, especialmente si corresponde a un documento administrativo. Si los usuarios finales no conocen estos documentos o fuentes de antemano la comprensión del alcance y la naturaleza de la información dependerá exclusivamente de su propia experiencia, por lo que una aproximación desde el campo de los archivos puede proporcionar una descripción pertinente del contexto en el que se ha creado la fuente original, la función a la cual corresponde, las competencias legales en las que se basa, quién los ha creado y por qué, puesto que esta información coincide en buena parte con la descripción archivística normalizada.

La disponibilidad de información contextual para cada fuente de información permite a los usuarios finales evaluar su exactitud o incluso cuestionar su credibilidad. Una descripción exacta del contexto de producción puede garantizar la autenticidad de la fuente de la cual se ha extraído el conjunto de datos.

Además, la misma metodología aplicada a la descripción del contexto en el cual se ha efectuado la extracción de datos contribuye a garantizar que no han sufrido ninguna manipulación que podría alterar su identidad, integridad y exactitud, facilitando así la disociación automática posterior

En general, la función transversal de la gestión de documentos proporciona a los profesionales una importante función de intermediación que permite identificar interlocutores adecuados en las unidades administrativas y, consecuentemente, ser facilitadores de soluciones a problemas específicos.

9. UNA DESCRIPCIÓN PRECISA DEL CONTEXTO DE CREACIÓN DE LOS DATOS PERMITE A LOS USUARIOS FINALES ENTENDERLOS MEJOR, ADEMÁS DE CONTRIBUIR AL FORTALECIMIENTO DE SU AUTENTICIDAD.

La mayoría de los proyectos open data describen en detalle los conjuntos de datos publicados desde el punto de vista de la estructura de datos, pero la información sobre el contenido suele ser muy limitada. Por otra parte, la fuente original ocasionalmente es descrita más allá de la inclusión del título, especialmente si corresponde a un documento administrativo. Si los usuarios finales no conocen estos documentos o fuentes de antemano la comprensión del alcance y la naturaleza de la información dependerá exclusivamente de su propia experiencia, por lo que una aproximación desde el campo de los archivos puede proporcionar una descripción pertinente del contexto en el que se ha creado la fuente original, la función a la cual corresponde, las competencias legales en las que se basa, quién los ha creado y por qué, puesto que esta información coincide en buena parte con la descripción archivística normalizada.

La disponibilidad de información contextual para cada fuente de información permite a los usuarios finales evaluar su exactitud o incluso cuestionar su credibilidad. Una descripción exacta del contexto de producción puede garantizar la autenticidad de la fuente de la cual se ha extraído el conjunto de datos. Además, la misma metodología aplicada a la descripción del contexto en el cual se ha efectuado la extracción de datos contribuye a garantizar que no han sufrido ninguna manipulación que podría alterar su identidad, integridad y exactitud.

**DEL TIEMPO
A LA LETRA**