

Los conjuntos de datos de la Corte Suprema de Justicia de la Nación. A propósito de las estadísticas y los *datasets* como forma de justicia abierta

por HELGA MARÍA LELL^(*) y RAMIRO ESCUDERO GIMÉNEZ^(**)

Sumario: I. INTRODUCCIÓN. – II. LOS CONJUNTOS DE DATOS DE LA CORTE SUPREMA. – III. DATOS Y CONJUNTOS DE DATOS. – IV. LOS FORMATOS DE PRESENTACIÓN DE LOS DATOS. – V. LA INTEROPERABILIDAD DE LOS DATOS. – VI. LA REUTILIZACIÓN DE LOS DATOS. – VII. CONSIDERACIONES FINALES.

I. Introducción

El 19 de marzo de 2024, la Corte Suprema de Justicia de la Nación (CSJN) decidió, mediante la Acordada 6/24⁽¹⁾, la creación de una Oficina de Estadísticas⁽²⁾. Las razones para ello son diversas: la enorme producción de datos sobre los trámites que este organismo lleva a cabo, la necesidad de sistematizar la información, la especificidad del procesamiento de datos, etc. Esta acordada nos conduce a pensar en un tema relacionado: nos referimos a la publicación de datos primarios y de los formatos de la presentación de dichos datos para la reutilización de maneras óptimas.

¿A qué nos referimos al hablar de “publicación de datos primarios” y “al formato”? Antes de responder esta pregunta, pensemos que las estadísticas implican un procesamiento de los datos para producir información, detectar regularidades y tendencias, localizar hitos llamativos, entre otras cuestiones. Entonces, la nueva oficina nos presentará información que implica la existencia de datos procesados para que podamos visualizar los reportes elaborados. No obstante, toda elaboración conlleva procesar la información con determinados fines, esto es, para exhibir adecuadamente aquello que se quiere mostrar. Ahora bien, si otras personas tuvieran intereses distintos a los de la Corte, respecto de aquello que se quiere saber, ¿podrían dejarse los datos a disposición para que cualquier interesado pueda elaborar sus propias estadísticas o realizar sus propias interpretaciones?

NOTA DE REDACCIÓN: Sobre el tema ver, además, los siguientes trabajos publicados en *EL DERECHO*: *Apostillas sobre la denominada apelación implícita*, por JORGE W. PEYRANO, ED, 187-519; *Meditaciones sobre la apelación (El fundamento constitucional, legal y doctrinario de la prohibición de reformatio in peius)*, por JUAN JOSÉ AZPÉLQUETA, ED, 195-959; *Meditaciones sobre la Apelación (Solución de cuestiones, vencimiento y ‘devolución’ implícita)*, por JUAN JOSÉ AZPÉLQUETA, ED, 195-1068; *Meditaciones sobre la Apelación (La adecuación oficiosa de costas y honorarios por la Alzada)*, por JUAN JOSÉ AZPÉLQUETA, ED, 200-950; *El origen de la anomia*, por ALBERTO M. SÁNCHEZ, ED, 255-938; *La doble instancia como garantía convencional*, por TORIBIO ENRIQUE SOSA, ED, 257-780; *Reposición inadmisibles y apelación en subsidio*, por JULIO CHIAPPINI, ED, 283-928; *Efecto adhesivo de los recursos en el Código Procesal Penal de la Nación*, por CARLOS ENRIQUE LERA, *Revista de Derecho Penal*, mayo 2021 - Número 5; *La implementación de la Apelación Horizontal a partir del precedente “Scalione” de la Corte Suprema de Justicia de Santa Fe*, por JONATAN ERIC PATERSON, *Revista de Derecho Penal*, junio 2023 - Número 6; *La Corte de Lorenzetti*, por ALFONSO SANTIAGO, EDCO, 2018-680; *La difícil formación de consensos en la Corte Suprema. Un análisis de algunos casos de los últimos cinco años de su jurisprudencia*, por JORGE ALEJANDRO AMAYA, *El Derecho Constitucional*, noviembre 2021 - Número 11; *Reflexiones para una reforma de la Corte Suprema. Análisis de los modelos de gestión en el derecho argentino y estadounidense*, por BAUTISTA CAÑÓN, *El Derecho Constitucional*, julio 2023 - Número 7; *Un legado en tomos: la publicidad de sentencias de la Corte Suprema de Justicia de la Nación ayer y hoy*, por FLORENCIA RATTI MENDAÑA y SOFÍA CALDERONE, *El Derecho diario*, La Corte Suprema: 160 años de jurisprudencia; *Anomia, polarización y precedentes*, de Leandro J. Giannini, por FLORENCIA RATTI MENDAÑA, ED, 307. Todos los artículos citados pueden consultarse en www.elderechodigital.com.ar.

(*) Doctora en Derecho (Universidad Austral). Posdoctora (Universidad Nacional de Rosario), Magíster y Especialista en Estudios Sociales y Culturales (UNLPam). Magíster en Filosofía (UNQ). Abogada (UNLPam). Investigadora adjunta de Conicet y docente de los doctorados en Derecho de la UCA, UNLPam/UNS/UNiCen y Universidad Austral.

(**) Abogado (Universidad de Buenos Aires). Ha participado en diversos proyectos de investigación DECyT (UBA), PICTO-UCA y PCyT (UCA), que involucraron el armado de bases de datos y el procesamiento de datos.

(1) CSJN, Acordada N° 6/2024, 19 de marzo de 2024.

(2) Cabe aclarar que la Corte Suprema de Justicia de la Nación argentina publica estadísticas actualizadas en su sitio web: <https://www.csjn.gov.ar/transparencia/datos-estadisticos> (fecha de consulta 27/8/2024).

Entonces, aquí nos centramos no en las estadísticas, sino en el paso previo: en la posibilidad de publicar los datos primarios y procesados, organizados en formatos adecuados para su reutilización por parte del mismo organismo, de otros tribunales o instituciones públicas y privadas, por académicos y científicos y por la ciudadanía.

Para retornar a la acordada, leamos el considerando 3°: “Que la Corte Suprema recibe, procesa, y produce una gran cantidad de datos sobre el esencial servicio de justicia que es llamada a prestar, sobre las materias que presentan litigiosidad en la sociedad, y sobre el accionar de las partes, letrados, tribunales inferiores y sobre su propio funcionamiento, tanto como tribunal supremo de justicia y como cabeza del Poder Judicial de la Nación”. Sin duda, la publicación de estadísticas sobre estas cuestiones es altamente valiosa, pero, si además de ello, la Corte Suprema publicase los datos que respaldan aquellos resultados de manera que puedan ser reutilizables, entonces, nos encontraríamos ante una multiplicación exponencial del valor y utilidad de ellos. No solo sería útil una Oficina de Estadísticas, sino que la labor de los científicos, académicos y otros interesados (como otros tribunales, el periodismo, ONG, colegios profesionales, etc.) en la temática podrían realizar aportes, así como identificar tendencias y realizar diferentes análisis estadísticos en el marco de una investigación cuantitativa.

Esta propuesta de publicación de datos se enmarca en la idea de justicia abierta, que se caracteriza, según Mercado, en un contexto global de transformación del vínculo entre el Estado y las personas, por “brindar a estas últimas un mejor entendimiento de la justicia (esto es, su funcionamiento, su apego a la ley y, en definitiva, el rol que cumple como reguladora de los intercambios sociales) contribuye a fortalecer el Estado de derecho, acrecentar la legitimidad de las decisiones que produce y transformar el servicio de justicia en uno más centrado en las personas y la satisfacción de sus necesidades legales”⁽³⁾.

Ahora bien, aunque aquí partimos de la acordada de la Corte Suprema y brindaremos ejemplos de la publicación de datos primarios del máximo tribunal, la propuesta es que el objetivo de hacer accesibles los conjuntos de datos sea compartido por todos los organismos judiciales, sea como forma de transparencia o como forma de construir conocimiento sobre el quehacer de las instituciones que imparten justicia de manera colaborativa.

II. Los conjuntos de datos de la Corte Suprema

La Corte Suprema de Justicia de la Nación posee un espacio específico para la publicación de datos en la sección “Datos abiertos”⁽⁴⁾. Allí se explica que se trata de una herramienta que tiene el objetivo “de difundir información pública y ponerla al alcance de todos, sin restricciones, en formatos digitales que posibilitan su reutilización y redistribución”⁽⁵⁾. Ahora bien, al ingresar a la página, podemos ver que publica distintos registros en formatos XLS, CSV y PDF. A la fecha, existen 252 *datasets* publicados, 113 corresponden al grupo “Presupuesto”, 112 a “Administración”, 7 a “Oficina de Violencia Doméstica”, 7 de “Base General de Datos de Niños, Niñas y Adolescentes”, 6 a “Oficina de la Mujer”, 5 a “Sentencias”, 1 a “Personal” y 1 a “Dependencias”. Puede verse que la

(3) MERCADO, Julio Gabriel, “Justicia Abierta y participación pública para un enfoque de justicia centrada en las personas”, *Reflexión política*, Vol. 25, N° 52, 2023, pp. 40-48, la cita se encuentra en la p. 42. Sobre el concepto de justicia abierta y qué parámetros hacen a su medición, véase SÁNCHEZ GONZÁLEZ, José Juan, “Transparentando los poderes públicos: Gobierno abierto, parlamento abierto y justicia abierta”, *Revista Venezolana de Gerencia*, Vol. 23, N° 81, 2018; JIMÉNEZ-GÓMEZ, Carlos, *Justicia abierta: transparencia y proximidad de la Justicia en el actual contexto de Open Government*, Centro de Estudios Jurídicos y Formación Especializada del Departamento de Justicia, 2014.

(4) La página web respectiva es: <https://datos.csjn.gov.ar/dataset> (fecha de consulta 26/8/2024).

(5) CSJN, “Qué es Datos abiertos”, en *Gobierno Abierto Judicial*: <https://datos.csjn.gov.ar/help>

mayoría se centran en cuestiones financieras y administrativas. En lo que aquí nos interesa, veamos el ítem que incluye información sobre las sentencias.

En el grupo “Sentencias”⁽⁶⁾ se incluyen *datasets* anuales desde 2012 hasta 2016. Cada conjunto de datos contiene, en realidad, escasa información sobre las causas y las sentencias. Más bien, solo se incluyen datos con una extensión de una carilla que resumen cuántas causas se han tratado y resuelto por secretaría. Entonces, la categoría se centra, principalmente, en cuánto ha hecho la Corte en cada año, pero no tanto en qué ha hecho como insumo para investigaciones, estudios o análisis de diversa índole. Cuando decimos “qué ha hecho” nos referimos a qué características tienen las sentencias como textos, qué temas se resolvieron, cómo se resolvieron, bajo qué acción llegó la causa a la Corte, etc. Un ejemplo de estudio de este tipo, sobre la redacción de las sentencias, fue realizado por uno de los actuales jueces de la Corte, Horacio Rosatti, en su tesis doctoral en Historia⁽⁷⁾. Allí compara la extensión de las sentencias medida en cantidad de palabras, la extensión de las oraciones, la comparación de la extensión de los fallos por materia, por la repercusión del propio fallo, el índice de claridad, la evolución de las citaciones, qué tipo de material se cita, etc. Es decir, en este estudio, la fuente primaria fueron los textos de las sentencias. Entonces, un conjunto de datos sobre las sentencias debería, por ejemplo, incluir los textos, las materias, fechas, resultados, etc., de manera tal que cualquier persona pueda luego procesarlos conforme a los interrogantes que tenga en mente⁽⁸⁾.

Además del mencionado trabajo de Rosatti que se focaliza en el análisis textual con diferentes parámetros para evaluar la claridad de las sentencias, pero también para caracterizar el quehacer de la Corte ante determinados casos, se destacan otros trabajos. Uno de ellos, también sobre la Corte Suprema, es la labor de Leandro J. Giannini que ha estudiado las decisiones dictadas y causas resueltas por este órgano en los años 2020 y 2021 tomando como parámetros las vías de acceso, materias, duración del proceso, origen de las causas apeladas, tasa de admisión y de revocación, técnicas de decisión (extensión, remisión a precedentes o al dictamen del Ministerio Público, cohesión, unanimidad, mayorías, disidencias, etc.), la participación de los jueces en las mayorías o minorías, cuántas causas se resolvieron de manera unánime, entre otros⁽⁹⁾. Desde otra perspectiva y ya no sobre las sentencias, sino sobre los discursos de apertura de cada año judicial, Macbeth, Razumiejczyk y Fernández analizaron el concepto de derecho subyacente en ellos⁽¹⁰⁾.

Estos son tan solo tres ejemplos de análisis factibles sobre distintos intereses que conducen a explorar y procesar datos de distintas maneras⁽¹¹⁾. Los datos requeridos son

(6) La información puede encontrarse aquí: <https://datos.csn.gov.ar/dataset?groups=sentencias> (fecha de consulta 26/8/2024).

(7) Cfr. ROSATTI, Horacio, *La palabra de la Corte Suprema. Cómo funciona, piensa y habla (y algunas ideas para debatir su futuro)*, Buenos Aires, Siglo XXI editores, 2022.

(8) Sobre los datos que son susceptibles de ser publicados para una justicia abierta, véase ELENA, Sandra; MERCADO, Julio Gabriel, “Justicia abierta: una aproximación teórica”, *Justicia abierta: aportes para una agenda en construcción*, Ediciones SAJ, 2018, pp. 117-42.

(9) Véase GIANNINI, Leandro J., “La producción jurisdiccional de la Corte Suprema de Justicia de la Nación (2020-2021)”, *Revista de Derecho Procesal*, 2022-2, pp. 489-571; GIANNINI, Leandro J., “La producción jurisdiccional de la Corte Suprema de Justicia de la Nación en su actual integración”, *Anales de la Facultad de Ciencias Jurídicas y Sociales de la Universidad Nacional de La Plata*, N° 48, 2018, pp. 1208-1259 y GIANNINI, Leandro J., “La producción jurisdiccional de la Corte Suprema de Justicia de la Nación [2018]”, *Anales de la Facultad de Ciencias Jurídicas y Sociales de la Universidad Nacional de La Plata*, N° 50, 2020, pp. 791-832. Cabe aclarar que este autor también ha realizado otra investigación de la misma índole sobre la justicia de la provincia de Buenos Aires (véase, GIANNINI, Leandro J., “Diagnóstico sobre el estado de la justicia civil. Resultados provisionales de una evaluación empírica en la Provincia de Buenos Aires”, *Anales de la Facultad de Ciencias Jurídicas y Sociales de la Universidad Nacional de La Plata*, N° 49, 2019, pp. 495-520).

(10) Véase MACBETH, Guillermo; RAZUMIEJCZYK, Eugenia y FERNÁNDEZ, Humberto, “Análisis lingüístico-cognitivo del derecho como sistema de reglas en discursos de la Corte Suprema de Justicia argentina”, 1° Congreso Internacional de Ciencias Humanas. Humanidades entre pasado y futuro. Escuela de Humanidades, Universidad Nacional de San Martín, Gral. San Martín, 2019.

(11) Existen muchos trabajos académicos que abordan las estadísticas de las cortes tanto supremas como inferiores en otros Estados. Por razones de extensión, aquí no nos detendremos en ellas, aunque sí quisiéramos hacer referencia. A título ejemplificativo, podemos mencionar los siguientes textos: LANDES, William, *An Economic Analysis of the Courts*. *The Journal of Law and Economics*, Vol. 14, N° 1, 1971, pp. 61-107; HAMANN, Hanjo. “The German Federal Courts Dataset

diversos, las finalidades son múltiples y las conclusiones también son variadas y enriquecedoras.

III. Datos y conjuntos de datos

Un dato primario es aquel que se recolecta de primera mano, esto es, de manera directa de la fuente original y que no ha sido procesado o interpretado previamente. Un dato procesado, en cambio, es aquel que ha atravesado algún proceso de tratamiento (digitalización, anonimización, etc.). Un conjunto de datos es toda aquella información colectada, observada o creada con el propósito de generar resultados originales⁽¹²⁾. La relevancia de estos conjuntos de datos es que, una vez explorados y procesados (por ejemplo, transformados en una matriz), permiten validar un descubrimiento, y establecer relaciones entre propiedades fundándolas en algo más que impresiones, así como cuantificar la fuerza de estas relaciones⁽¹³⁾. De esta manera, todo dato es un insumo potencialmente relevante para una investigación o para el uso con determinados fines. Para ello, los datos deben ser localizables, conservables y, sobre todo, reutilizables. Asimismo, se debe asegurar su calidad e integridad.

Los conjuntos de datos deben ser precisos, exactos y descriptivos. Deben contener variables e información necesaria para su análisis, además de que deben ser actualizados. Deben ser consistentes, lo cual implica que deben ser coherentes y veraces. Deben ser representativos o completos, es decir, deben contar con un volumen suficiente, en caso de tratarse de una muestra, o bien deben contemplar la totalidad del universo. Deben ser legibles, lo que implica que se encuentren estructurados de manera comprensible. Deben respetar el anonimato y la privacidad en caso de existir datos sensibles. Finalmente, deben ser accesibles, o sea, estar en un formato disponible para su reutilización⁽¹⁴⁾.

IV. Los formatos de presentación de los datos

Los conjuntos de datos pueden ser presentados en distintos formatos según su naturaleza y propósitos de usos estimados. Algunos que, según los datos que podrían proporcionar los tribunales, parecen más adecuados, son⁽¹⁵⁾:

- Formato CSV (Valores separados por comas), en el cual los datos aparecen en forma de texto, cada línea representa un registro y las columnas están separadas por

1950-2019: From Paper Archives to Linked Open Data”, *Journal of Empirical Legal Studies*, N° 16, 2019, pp. 671-688; ASH, Elliott; GOEL, Naman; LI, Nianyun; MARANGON, Claudia & SUN, Peiyao, “Curated Large Dataset of Criminal Cases from Wisconsin Circuit Courts”, Oh, A.; T. NAUMANN; A. GLOBERSON; K. SAENKO; M. HARDT and S. LEVINE (eds.), *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2023, pp. 12626-12643; NGIGE, O. C., AYANKOYA, F. Y., BALOGUN, J. A., ONUIRI, E., AGBONKHESE, C., SANUSI, F. A., “A dataset for predicting Supreme Court judgments in Nigeria”, *Data in Brief*, Vol. 50, 2023, disponible en: <https://doi.org/10.17632/ky6zfyf669.1> (fecha de consulta 27/8/2024); BOYD, Christina, Kim, Pauline; SCHLANGER, Margot, “Mapping the Iceberg: The Impact of Data Sources on the Study of District Courts”, *Journal of Empirical Legal Studies*, Vol. 17, N° 3, 2020, pp. 466-492; CUI, Junyun; SHEN, Xiaoyu; WEN, Shaochun, “A Survey on Legal Judgment Prediction: Datasets, Metrics, Models and Challenges”, *IEEE Access*, Vol. 11, 2023, pp. 102050-102071; MARKOVI, Marko, GOSTOJI, Stephan, “Open Judicial Data: A Comparative Analysis”, *Social Science Computer Review*, Vol. 38, N° 3, 2020, pp. 295-314; WEINSHALL, Karen; EPSTEIN, Lee, “Developing High Quality Data Infrastructure for Legal Analytics: Introducing the Israeli Supreme Court Database”, *Journal of Empirical Legal Studies*, Vol. 17, N° 2, 2020, pp. 416-434; DELEN, Dursun; ZOLBANIN, Hamed; CROSBY, Durand; WRIGHT, David, “To imprison or not to imprison: an analytics model for drug courts”, *Annals of Operations Research*, Vol. 303, N° 1-2, 2021, pp. 101-124; PINA-SÁNCHEZ, José; GRECH, Diana; BRUNTON-SMITH, Ian; FERROPOULOS, Dimitrios, “Exploring the origin of sentencing disparities in the Crown Court: Using text mining techniques to differentiate between court and judge disparities”, *Social Science Research*, Vol. 84, disponible en: <https://doi.org/10.1016/j.ssresearch.2019.102343>; MAHARI, Robert; STAMMBACH, Dominik; ASH, Elliott; PENTLAND, Alex, “LePaRD: A Large-Scale Dataset of Judicial Citations to Precedent”. KU, Lun-Wei; MARTINS, Andre; SRIKUMAR, Vivek (eds.), *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, Vol 1, Aug. 2024, Bangkok, Thailand. Association for Computational Linguistics, pp. 9863-9877.

(12) Cfr. Grupo de Trabajo de “Depósito y gestión de datos en acceso abierto” del proyecto RECOLECTA. *La conservación y reutilización de los datos científicos en España. Informe del grupo de trabajo de buenas prácticas*, Fundación Española para la Ciencia y la Tecnología, FECYT, 2012.

(13) Cfr. MARRADI, Alberto; ARCHENTI, Nélida & PIOVANI, Juan Ignacio, *Metadología de las Ciencias Sociales*. Emecé, 2007, p. 24.

(14) Cfr. SCOTT, Mark; STARK, Isobel, BYATT, Dorothy & COX, Simon, *Introducing research data*, Stanford University, 2016.

(15) Cfr. OPEN KNOWLEDGE INTERNATIONAL, “Formato de Archivos”. *Open Data Handbook*, disponible en: <https://opendatahandbook.org/guide/es/appendices/file-formats/> (fecha de consulta 26/8/2024).

comas (o el valor delimitador que se determine). Este formato es muy común porque es compatible con la mayoría de los programas de hojas de cálculo y herramientas de análisis de datos y son fácilmente legibles por lenguajes de programación. Si bien puede ser muy útil para datos simples que requieran diferentes categorías, puede ser un formato un tanto engorroso para textos extensos y complejos.

- Formato JSON (Notación de Objetos de JavaScript), que es un formato ligero basado en texto, útil para estructuras de datos complejas. Su principal ventaja es su estructura y formato, que permite presentar los datos de forma organizada, y con una sintaxis simple y concisa. Asimismo, es agnóstico respecto del lenguaje de programación, lo que significa que podrá usarse el lenguaje que se desee para analizar los datos⁽¹⁶⁾. Es el formato más extendido para el intercambio de datos en la web.

- Formato XML (Lenguaje de marcado extensible [*Extensible Markup Language*]), que es un formato muy interesante para la edición de textos y representar datos jerárquicos en ellos. La desventaja que posee es que es necesario conocer el sistema de marcado, además de realizar la respectiva edición. Por otro lado, resulta de procesamiento más dificultoso en términos de performance que otros formatos (por ejemplo, JSON)⁽¹⁷⁾ por los lenguajes de programación, por lo que actualmente su uso se encuentra en declive.

- SQL (Lenguaje de consulta estructurada), que, si bien no es un formato de archivo, es un lenguaje utilizado para interactuar con bases de datos relacionales. Posee la fortaleza de ser altamente útil para el manejo de bases de datos. Estas bases de datos relacionales permiten la organización de datos de manera metódica, organizándolos en tablas que contienen entradas, pudiendo establecerse relaciones entre ellas de forma simple, y recuperar las mismas mediante el lenguaje SQL.

- Formato TXT (Texto plano), que es, quizás, el formato más sencillo de todos. Implica archivos en los que los datos están separados por delimitadores como tabulaciones o espacios. Es legible por todos los sistemas y puede ser útil para datos no estructurados.

- Formato XLS/XLSX (hoja de cálculo de Microsoft Excel), si bien no es un formato propio de bases de datos, sí permite operar con ellos fácilmente, además de poseer la ventaja de poder utilizar un programa conocido. Es, asimismo, procesable por lenguajes de programación con librerías como OpenPyXL o Pandas (para el lenguaje de programación Python) o por suites de análisis de datos como SPSS o SOFstats.

- Formato DOC/DOCX (documento de Microsoft Word), si bien su finalidad no es estrictamente almacenar datos textuales, sino la edición de textos, es un formato compatible con la mayoría de los procesadores de texto. Resulta programáticamente analizable el texto (aunque con algunas dificultades, por el propio formato, diferente del texto plano) y es posible utilizar, por ejemplo, *software* de análisis lingüístico (como LIWC) en este tipo de archivos.

Tal como se mencionó, la Corte Suprema, en el espacio dedicado a las bases de datos, publica archivos en formatos CSV, XLS y PDF. Los primeros dos resultan adecuados para una reutilización simple, mientras que el último no resulta idóneo para la lectura automática. En primer lugar, el PDF (formato de documento portátil) no resulta fácilmente editable. Los tomos más antiguos no poseen un OCR que permita leer automáticamente el texto⁽¹⁸⁾. En segundo término, la lectura por un *script* no puede llevarse a cabo identificando variables, por ejemplo, por filas y columnas, como sí puede hacerse con un documento en formato CSV o un XLS. Es posible, utilizando diversas librerías, como Tabula-py, convertir las tablas que existan

en el archivo en información procesable, pero este es un proceso muchas veces inexacto y que requiere tiempo y procesamiento manual de la información para evitar errores en la misma.

Cabe agregar que, respecto de las sentencias, resulta interesante que en los últimos años se ha agregado cada vez que se cargan las sentencias del acuerdo un segundo formato de publicación, y junto con el PDF, es posible acceder al documento en formato DOC. Este segundo formato, reciente para la publicación y que solo se encuentra en los fallos de los últimos años, permite procesar el texto para, por ejemplo, convertirlo en un TXT, o también para, a fines de una investigación en particular, seleccionar fragmentos a analizar o bien para realizar un marcado XML con los editores correspondientes.

V. La interoperabilidad de los datos

Una de las principales ventajas de la publicación de datos primarios y de conjuntos de ellos en formatos adecuados es la interoperabilidad que permite que diferentes aplicaciones puedan compartir, acceder y utilizar información con independencia de sus plataformas o formatos. Entre las ventajas de la interoperabilidad, podemos mencionar que permite que distintas instituciones e incluso particulares puedan utilizar los datos de manera más fluida. Esto conduce a maneras más eficientes de intercambiar datos, a reducir errores y contradicciones y, por supuesto, a optimizar tiempos y recursos en la producción de la información. Por otro lado, en aquellos casos en los que un organismo o individuo tomara datos de diferentes fuentes para comparar puede permitir obtener resultados más innovadores⁽¹⁹⁾.

La interoperabilidad podría lograrse fácilmente a través de una API. Esta es una Interfaz de Programación de Aplicaciones que es un conjunto de protocolos que permite que diferentes aplicaciones se comuniquen entre sí. Las API funcionan como un punto intermediario para que una *software* pueda acceder a los datos de otro. Esto es, se solicita, envía o recibe datos de otro sistema de manera estructurada⁽²⁰⁾. La Corte Suprema informa en su página web que es posible ingresar a través de la API⁽²¹⁾. Sin embargo, no se encuentra documentada actualmente: pudimos comprobar que el *link* que refiere a la documentación de la API lleva a una página inexistente, lo que dificulta el acceso a esta herramienta que la Corte Suprema pone a disposición. En este sentido, no se ha podido relevar, por esta razón, el tipo de información contenida, y tan solo se ha constatado su existencia⁽²²⁾.

VI. La reutilización de los datos

La reutilización de datos primarios es relevante en términos de construcción colaborativa de conocimiento, la posibilidad de no duplicar datos o impedir la recolección de datos contradictorios y hace a la eficiencia del uso de recursos disponibles.

Pensemos en la cantidad de tiempo que insume la recolección de datos primarios. El hecho de que alguien que ya ha hecho esta labor los haga públicos permite que otras personas puedan utilizarlos. Si, además, se tiene en cuenta que el sujeto que dispone de los datos es una entidad pública, las razones para que publique estos datos se acrecientan. Así, por ejemplo, si los datos están disponibles, los investigadores, académicos, periodistas, etc., pueden tomar aquella información abierta y pública y concentrarse en el análisis sin tener que repetir el proceso de recolección. Desde otra mirada, el hecho de el productor de los datos que sea un organismo público otorga un respaldo de veracidad, además de que esta información pública

(16) Cfr. ECMA International. *Standard ECMA-404: The JSON Data Interchange Syntax*. <https://ecma-international.org/publications-and-standards/standards/ecma-404/>

(17) Cfr. ZUNKE, Saurabh & D'SOZA, Veronica, "JSON vs XML: A comparative Performance Analysis of Data Exchange Formats", *IJCSN International Journal of Computer Science and Network*, Vol. 3, N° 4, 2014, pp. 257-261.

(18) Sobre la publicación de los tomos históricos de la Corte Suprema, véase RATTI MENDAÑA, Florencia, CALDERONE, Sofía, "Un legado en tomos: la publicidad de sentencias de la Corte Suprema de Justicia de la Nación ayer y hoy". *El Derecho. Suplemento Especial La Corte Suprema: 160 años de Jurisprudencia*, 14 de noviembre de 2023; disponible en: <https://repositorio.uca.edu.ar/handle/123456789/17693> (fecha de consulta 27/8/2024).

(19) Cfr. CUBIDES, Jhon Francined HERRERA; Alonso GAONA GARCÍA, Paulo y SÁNCHEZ ALONSO, Salvador, "Linked Data: qué sucede con la heterogeneidad y la interoperabilidad", *Scientia et Technica*, Vol. 23, N° 2, 2018, pp. 230-240.

(20) Cfr. ROSAS TORO, María Alejandra; HERRERA, Hesiquio; VELASCO SALINAS, Jaime; SUZETTE, Leslie y MARTÍNEZ CHÁVEZ, Juan Manuel, "Construyendo una web API con la arquitectura en capas", *Multidisciplinas de la ingeniería*, Vol. 11, N° 18, 2023, pp. 14-25; y PLAZA ESTÉVEZ, Sheila; RAMÍREZ LAMELA, Nerea y ACOSTA MORALES, Carmen, "API de servicios web orientados a accesibilidad", trabajo de fin de grado en Ingeniería Informática, Universidad Complutense de Madrid, Facultad de Informática, 2016.

(21) Véase: <https://datos.csjn.gov.ar/dataset?tags=csjn&organizacion=csjn> (fecha de consulta 26/8/2024).

(22) Se puede acceder directamente a la API mediante la siguiente URL: <https://datos.csjn.gov.ar/api/3> (fecha de consulta 26/8/2024).

es accesible para cualquier persona que quiera comprobar dichos datos.

A ello cabe agregar que el hecho que otras personas puedan explorar, procesar e interpretar los datos es un puntapié para nuevas perspectivas y, por ende, la posibilidad de descubrir patrones y relaciones no detectadas por otros.

VII. Consideraciones finales

La publicación de los datos primarios y de los conjuntos de datos en formatos accesibles tiene el potencial de enriquecer la información pública y de expandir los estudios sobre el quehacer de la Corte Suprema. Así, la Oficina de Estadísticas es solo un paso para poder mostrar lo que el mismo organismo quiere exhibir, esto es, implica información procesada conforme con fines. Pero los intereses pueden ser múltiples.

Ahora bien, entre las ventajas de publicar los datos primarios se destaca justamente la antedicha: es posible obtener muchos más resultados que los que puede procesar la oficina creada por la Acordada 6/24. Cada interesado, a partir del acceso a los datos, puede pensar distintos parámetros, nuevas preguntas y diferentes proyecciones. Así, la publicación de los datos para su reutilización tiene un poder expansivo y multiplicador.

Por el lado de las desventajas, cabe mencionar que la publicación de los datos también requiere de un espacio de almacenamiento y de capacidad de respuesta ante los requerimientos. Por supuesto, eso insume costos económicos y requiere de infraestructura tanto humana como tecnológica.

Para finalizar, podemos señalar que la pregunta sobre si es conveniente o no publicar los datos primarios es quizás un tanto obsoleta. La tendencia actual es que los organismos públicos pongan a disposición *datasets* de manera abierta, no solo en Argentina sino también en el mundo. Aun si un organismo no se dedica a hacerlo, cualquier interesado podría recopilar esos datos y publicarlos, ya que la información es pública. Por supuesto que el trabajo es mucho más laborioso y quizás casi manual respecto del organismo que posee esos datos de primera mano. También se corren otros riesgos, como la incoherencia de los datos recolectados por distintos individuos. En tal sentido, la publicación de los datos primarios por parte de entidades públicas, entre ellas, la Corte Suprema, resulta ventajoso y de suma utilidad.

A modo de síntesis, podemos señalar que la creación de la Oficina de Estadísticas es una novedad para celebrar, que se suma a la información existente. Desde luego que también es relevante continuar con la labor de completar el acervo de *datasets* de la Corte Suprema más allá de la información administrativa.

VOCES: CÁMARAS DE APELACIONES - JURISPRUDENCIA - SENTENCIA - PODER JUDICIAL - INTERPRETACIÓN DE LA LEY - LEY - CONSTITUCIÓN NACIONAL - FILOSOFÍA DEL DERECHO - PODER JUDICIAL - CORTE SUPREMA DE LA NACIÓN - RECURSO DE INAPLICABILIDAD DE LEY - DERECHO PROCESAL - PROCESO JUDICIAL - RECURSOS PROCESALES - DEMANDA - INFORMÁTICA - TECNOLOGÍA - DERECHO ADMINISTRATIVO