

Vancouver Statement on Collections-as-Data

Padilla, Thomas, Scates Kettler, Hannah, Varner, Stewart, & Shorish, Yasmeeen. (2023). Vancouver Statement on Collections as Data (ES). Zenodo. <https://doi.org/10.5281/zenodo.8341570>.

Última actualización: 2023

Precedido por: *Santa Barbara Statement on Collections-as-data*¹, 2017

Desde que se publicó la declaración *Santa Barbara Statement on Collections-as-data*² (2017) se ha producido un crecimiento en la participación en colecciones usadas como fuentes de datos internacionalmente. Las instituciones, tanto grandes como pequeñas, o de forma individual o colectiva, han invertido en el desarrollo, la facilitación de acceso y el respaldo del uso computacional responsable de colecciones usadas como fuentes de datos. Por ello, es necesaria una declaración actualizada a la luz de una mayor implementación comunitaria de colecciones usadas como fuentes de datos en el contexto de un entorno de datos cada vez más complejo.

La declaración *Vancouver Statement* sugiere un grupo de principios para cavilar las cuestiones que el trabajo de Collections-as-data genera, como parte de un esfuerzo de expansión global, interprofesional e interdisciplinar para empoderar la memoria, el conocimiento y a los responsables de la gobernanza de los datos (facultativos y académicos)³, cuyo objetivo es dar soporte al desarrollo responsable y el uso computacional de colecciones usadas como fuentes de datos. Este rol de responsabilidad de gobernanza solo crece en importancia a medida que las aplicaciones de inteligencia artificial, entrenadas con enormes cantidades de datos, (incluidas las colecciones usadas como fuentes de datos) impactan en nuestras vidas cada vez de forma más penetrante.

El concepto de Collections-as-data surge de un momento particular en la historia de galerías, bibliotecas, archivos y museos. Durante décadas, facultativos y académicos han ido creando colecciones digitales. Simultáneamente, una amplia y diversa gama de grupos de usuarios (p. ej., académicos, artistas, legisladores, empresarios, reporteros y facultativos GLAM [de galerías, bibliotecas, archivos y museos]) han recurrido a medios computacionales para obtener información de las colecciones a escala. Este trabajo usa una gran variedad de métodos que incluyen, entre otros, la minería de textos, la visión por computadora, el aprendizaje automático, la inteligencia artificial, la visualización de datos, el mapeo, el análisis de imágenes, el análisis de audio y el análisis de redes. En todo el mundo, se han desarrollado iniciativas para respaldar el desarrollo responsable y el uso computacional de las colecciones. No obstante, todavía queda trabajo para iniciar, respaldar y socializar enfoques que exploren una amplia gama de recursos y prioridades estratégicas locales, regionales, nacionales e internacionales.

¹ <https://zenodo.org/record/3066209>

² En este documento se utiliza la apariencia con guión “Colecciones como datos” para referirse al concepto, la comunidad y el área de práctica involucrados en respaldar el uso computacional responsable de las colecciones como datos.

³ Responsable: una persona que facilita y promueve el acceso y el cuidado a largo plazo de las colecciones.

Aunque los detalles de cómo desarrollar, ofrecer acceso y respaldar el uso de colecciones usadas como fuentes de datos pueden variar, cualquier material digital puede potencialmente ponerse a disposición como datos susceptibles de uso computacional. El uso y la reutilización de datos puede fomentarse con formatos adecuados con o sin licencia, en particular si los datos se ponen a disposición de formas muy diversas y el acceso se diseña para cumplir con las necesidades de la comunidad y tiene presente las cuestiones éticas.

El entendimiento y la atención a las cuestiones éticas son esenciales para Collections-as-data. Las iniciativas de Collections-as-data deben hacer que se tomen decisiones informadas sobre el acceso por parte de diversas prioridades comunitarias. Las posiciones anticipadas por los Principios CARE de la Alianza global de datos indígenas⁴ y el trabajo sobre la Soberanía de los datos maoríes⁵ son instructivos. Debe tenerse cuidado en el cumplimiento de los requisitos legales y las necesidades de accesibilidad, respetarse las normas y valores culturales de diversidad y tener en cuenta los riesgos para los grupos vulnerables, además de respetar los derechos y la autoridad de estas comunidades.

La escala de algunas colecciones puede la integridad (o ausencia de integridad) de las historias perceptibles de ser representadas. Los facultativos y académicos de memoria, conocimiento y datos deben ser conscientes de estas ausencias o distorsiones y trabajar para evitar su repetición. La documentación debería informarse según las prácticas de los principios documentales y de replicabilidad efectiva para garantizar que los usuarios dispongan de la información que necesitan para trabajar con las colecciones de forma responsable.

La declaración *Vancouver Statement* se creó para ser accesible tanto por aquellos para los que la idea de respaldar el desarrollo responsable y el uso computacional de las colecciones usadas como fuentes de datos es nueva, como por aquellos con mayor experiencia en enfoques establecidos desde hace tiempo acerca de las prácticas sobre galerías, bibliotecas, archivos y museos.

Las implicaciones de los principios descritos a continuación no tienen por qué ser abordadas de una sola vez. De hecho, puede que algunas sigan siendo muy ambiciosas durante algún tiempo, pues las personas comienzan este trabajo en colaboración con colegas con diferentes experiencias, valores y enfoques sobre el trabajo. La cuestión está en comenzar por algún sitio, valorar, ajustar y mejorar.

1. **El desarrollo de Collections-as-data respalda el uso computacional responsable de las colecciones digitales, tanto las digitalizadas como las nativas digitales.** Los responsables de los conocimientos, datos y memorias amplían el grupo de oportunidades de participación en las colecciones al concebir, crear y ofrecer acceso a las colecciones usadas como fuentes de datos.

⁴ Carroll, S.R., Garba, I., Figueroa-Rodríguez, O.L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J.D., Anderson, J. and Hudson, M., 2020. The CARE Principles for Indigenous Data Governance. *Data Science Journal*, 19(1), p.43. DOI: <https://doi.org/10.5334/dsj-2020-043>

⁵ <https://www.temanararaunga.maori.nz/>

2. **Los responsables de Collections-as-data apuestan por el trabajo contra las desigualdades históricas y contemporáneas representadas en la adquisición, el ámbito, la descripción, el acceso y el uso de datos.** Los compromisos de trabajar contra las desigualdades deben documentarse y ponerse a disposición pública, a excepción de los casos en los que dichos compromisos pongan en riesgo público a las comunidades. Esta documentación deberá incluir (cuando sea aplicable) todos los silencios o sesgos conocidos en las colecciones. Los detalles de cada compromiso variarán en cada comunidad a la que sirvan las colecciones, pero compartirán la causa común de buscar afrontar las necesidades de la vulnerabilidad. Los responsables podrán imponer barreras meditadas para su uso, incluida la abstención de poner una colección usada como fuente de datos a disposición pública. Los responsables respetarán los derechos y necesidades de las comunidades que creen contenido que constituya colecciones, aquellas que estén representadas en colecciones y las comunidades que las utilicen.
3. **Las colecciones usadas como fuentes de datos deben ser accesibles de forma generalizada, dentro de los límites de las expectativas éticas, legales y de la comunidad.** Los términos de uso de las colecciones usadas como fuentes de datos deben explicitarse, y deberá informarse en prácticas comunitarias como RightsStatements.org, los principios CARE, y licencias estándar como Creative Commons⁶, Open Data Commons⁷ y licencias Traditional Knowledge⁸. Los materiales deben ser compatibles con las diferentes herramientas y programas de accesibilidad, y deben estar disponibles en varios idiomas o en un formato que pueda ser traducido mediante traducción automática.
4. **Collections-as-data aprovechan el diseño participativo.** En lugar de presuponer las necesidades o de imaginar las comunidades de usuarios, los responsables deben colaborar con otros y trabajar juntos para reducir las barreras en el uso (cuando corresponda), pensar en principios de diseño universales y seguir valorando las necesidades de los usuarios a lo largo del tiempo. Si los recursos lo permiten, se recomiendan varios enfoques en cuanto al desarrollo y el acceso a los datos para facilitar así la interoperabilidad (como se describe en el Principio 6). Las personas deberán estar motivadas y alentadas para crear y compartir el resultado de este trabajo.
5. **Compartir documentación aporta contexto y ayuda a otros a encontrar el camino para llevar a cabo el trabajo.** El trabajo de Collections-as-data conlleva tomar decisiones sobre desarrollo, suministro de acceso y respaldo de uso. La documentación deberá explicar claramente cómo y por qué se han tomado dichas decisiones en el trabajo de Collections-as-data. Aunque ninguna documentación puede ser totalmente detallada, una documentación incompleta o en curso siempre es mejor que ninguna documentación en absoluto. Algunos ejemplos de documentación incluyen planes estratégicos, descripciones de trabajos, diagramas de metadatos comprensibles (tanto de humanos como de máquinas), fichas técnicas, flujos de trabajo, perfiles de aplicación, escrituras de donación y libros de códigos, con su traducción en diferentes idiomas. La elección del idioma debe reflejar las necesidades de los usuarios. La documentación

⁶ <https://creativecommons.org/about/ccllicenses/>

⁷ <https://opendatacommons.org/>

⁸ <https://localcontexts.org/licenses/traditional-knowledge-licenses/>

debe ser de acceso público y tenaz, teniendo en cuenta las consideraciones éticas y legales pertinentes.

6. **El desarrollo de Collections-as-data valora la interoperabilidad social y técnica.** La interoperabilidad conlleva el ajuste con los modelos (personal, equipos matriciales, mantenimiento y estrategias de innovación), estándares e infraestructura de la comunidad establecida o emergente y facilita la integración con una infraestructura centralizada, además de distribuida. Deben tenerse en cuenta los sesgos presentes en los estándares y la infraestructura. La interoperabilidad va más allá de lo técnico, hasta la interoperabilidad social; la capacidad de las personas de interactuar de forma efectiva y respaldar la infraestructura humana sostenible.
7. **Los responsables de Collections-as-data trabajan con total transparencia para mantener la integridad del acceso a largo plazo a las colecciones.** La integridad de Collections-as-data depende de las iniciativas para garantizar y documentar públicamente la integridad técnica de los datos, su procedencia y las relaciones presentes entre los responsables de las colecciones y las comunidades o los grupos de usuarios. Este proceso requiere que los responsables confirmen las ausencias, los sesgos y las áreas de incertidumbre. Las colecciones usadas como fuentes de datos transparentes deben incluir metadatos abiertos y sólidos (de acuerdo con el Principio 3) y deben tener un compromiso de preservación asociado a ellos.
8. **La progresión del trabajo de Collections-as-data depende de los compromisos organizativos con una infraestructura y unos procesos sostenibles.** El desarrollo continuo de las colecciones usadas como fuentes de datos puede afectar a los modelos de personal, los flujos de trabajo y la infraestructura técnica. Las entidades y las organizaciones deben respaldar el desarrollo de colaboraciones productivas y las integraciones entre los flujos de trabajo nuevos y los ya existentes, además de los cargos del personal y los modelos de servicio.
9. **El trabajo del desarrollo de colecciones usadas como fuentes de datos debe equilibrar los beneficios con la preocupación por el impacto climático.** El trabajo digital, por naturaleza, depende de centros de datos que utilizan grandes cantidades de energía y utilizan metales pesados y otros materiales perjudiciales para el medio ambiente. Siempre que sea posible, diseñe procesos que aprovechen los recursos computacionales que sean transparentes sobre las formas en las que afectan negativamente o no al medio ambiente.
10. **El trabajo del desarrollo de colecciones usadas como fuentes de datos debe equilibrar los beneficios con la preocupación por la explotación laboral.** La creación de colecciones usadas como fuentes de datos puede que haya sido (o sea) propiciada por alguna forma de explotación laboral. Los creadores deben confirmar los contextos laborales (siempre cuando se conozcan) mediante una documentación apropiada de los datos y la utilización de clasificaciones como CRediT⁹.
11. **Los responsables de Collections-as-data deben diseñar el acceso con una consideración adecuada del consumo de datos por parte de la inteligencia artificial u otras tecnologías.** Las colecciones usadas como fuentes de datos pueden utilizarse en conjuntos de entrenamiento para la inteligencia artificial y el desarrollo de productos. Los responsables deben ser conscientes

⁹ <https://credit.niso.org/>

de las implicaciones éticas, incluidas la propiedad intelectual y las reclamaciones sobre la soberanía de los datos, y deben desarrollar garantías adecuadas contra el uso indebido, la pérdida dañina de contexto y la amplificación de sesgos a través de estas tecnologías.