

Módulo 2: Bases de datos

[00:00:12] **Sandra Crucianelli** Hola a todos. Continuamos con el Módulo 2, que se centra esta semana en saber dónde están los datos, cómo podemos hacer para encontrar grandes volúmenes de información. En el último video les comenté las técnicas más sencillas de búsquedas que aplicamos a los principales buscadores, me estoy refiriendo a Google y a Bing, seguramente ustedes van a poder practicar búsquedas avanzadas en ambos buscadores. Pero también, como les comenté en el último video, hay un nivel de profundidad dentro de la web que es donde se alojan miles y miles de documentos en distintos formatos y están alojados en bases de datos. Por lo tanto, es una buena estrategia poder saber cuáles son esas principales bases de datos y cómo acceder a este tipo de documentos.

[00:01:04] **Sandra Crucianelli** Les voy a mostrar algunos de los recursos más interesantes que tenemos y que usamos generalmente en periodismo de datos. La primera herramienta que les presento es Archive.org. Es como la mayor biblioteca digital que hay en la web y que tiene alojados documentos de distinto tipo, no solamente libros, películas, software, música, sitios webs y mucho más en diferentes formatos. Deben explorar lo que hay aquí para cada uno de sus países y de acuerdo sea al interés que ustedes tengan. Pero también tienen esta herramienta que es la Way Back Machine. La Way Back Machine es una herramienta que les permite recuperar información de sitios web que ya no están en la web y de la cual esta herramienta ha tomado una captura de pantalla y allí ha quedado. Por ejemplo, si colocamos -- yo trabajo en Infobae -- si colocamos la raíz de un dominio sobre el cual queremos encontrar información, vamos a ver la progresión de ese sitio web en esta línea de tiempo, es decir, desde las primeras capturas de pantalla que quedaron registradas en este caso en el año 2002. Y una vez que ingresan al año en el que ustedes desean consultar qué decía ese sitio web -- aquí, por ejemplo, tenemos capturas de pantalla del 6 de junio de 2004 -- sólo tienen que hacer clic sobre los círculos que están marcados sobre el calendario y esperar un tiempito a que la base de datos le devuelva la información. A veces podría -- ahí, ahí está recuperando -- el resultado sería una homepage del año 2004, es decir de hace mucho tiempo. Y esta es la manera de recuperar información de sitios webs de hace muchos años ya que, como siempre decimos, todo lo que se publica en la web queda en la web y por más que se elimine. Vamos a suponer un sitio web que ya no está, esto pasa con sitios..., direcciones URL de sitios gubernamentales que pueden desaparecer. Sin embargo, ustedes los pueden recuperar usando la way back machine, la máquina del tiempo, de esta fabulosa herramienta que es Archive.org. Y se las dejo como inquietud para que prueben a ver qué encuentran en ellas.

[00:03:47] **Sandra Crucianelli** En cuanto a las bases de datos, hay una colección interesante. Yo he subido y he compartido estos recursos a través de una plataforma que se llama LiveBinders.com, ustedes pueden ingresar www.livebinders.com, un binder es como una suerte de bibliorato. Aquí hay sobre periodismo de datos, decenas y decenas de recursos sobre distintos temas que tienen que ver con el curso que nos ocupa hoy. No solamente sobre búsqueda de datos sino sobre narrativas basadas en datos, sobre los productos que elaboran los equipos de datos, las iniciativas de aperturas de datos que hay en distintos países del mundo, las herramientas de inteligencia artificial que podemos usar en periodismo de datos, todo lo que tienen que ver con seguridad digital, y también van a ver aquí un bibliorato sobre bases de datos, que es el tema que nos ocupa en esta semana. Lo que yo he hecho aquí, este bibliorato se navega a través de pestañas, las pestañas superiores en azul, que en la medida que ustedes las vayan abriendo van a ver

subpestañas, que están marcadas en gris, y cada una estas subpestañas tiene un link que vendría a ser como un repositorio de favoritos organizados temáticamente.

[00:05:08] **Sandra Crucianelli** Entonces el primer punto a tener en cuenta es averiguar qué datos abiertos hay de mi país en los sitios de gobierno abierto, es decir, estos operan como base de datos. Aquí estoy viendo el portal de datos abiertos de Chile. También tiene catálogos de datos abiertos Uruguay. Voy a abrir el de Brasil para que lo vean. El de Colombia ya lo hemos visto en la semana pasada. Aquí tenemos el de Uruguay. Aquí tenemos el de Chile. Aquí tenemos el de Brasil. Acá tenemos el de Paraguay. Aquí tenemos el sitio de datos abiertos de Estados Unidos que tiene mucha información sobre comercio exterior, básicamente todo el comercio exterior que hay entre los países de América Latina o de habla hispana y el gobierno norteamericano. Aquí tienen el sitio de gobierno abierto de Panamá. Y también voy a abrir el de Costa Rica para que vean que prácticamente casi todos los países de América Latina, bueno España por supuesto, hay una gran cantidad de países que se han unido en lo que se llama la Alianza de Gobierno Abierto. La Alianza de Gobierno Abierto, la Open Government Partnership, es una institución que se creó en el 2011, a la cual los principales países del mundo se han sumado. Aquí pueden listar los miembros nacionales, las naciones que están, que han hecho compromisos para abrir datos. Ahí pueden buscar. Vamos a suponer el caso de Chile. Yo voy a abrir Chile ahora aquí en este buscador por países, son los países que han hecho compromisos, han firmado acuerdos internacionales, han asumido compromisos para abrir datos públicos. Por lo tanto, la primer base de datos que deben conocer es la base de gobierno abierto del país en el que viven, ver qué hay. Y también un posible trabajo periodístico sería descargar estos recursos, como son los planes de acción que se ha comprometido cada país a llevar a cabo en los años pasados y en el presente, y ver si han prometido o no todo con lo que se cumplió. Esa es una buena manera de empezar. Y después recorrerlos, estos portales para ver, qué es lo que hay en cada uno de ellos. Siempre es una buena manera de comenzar explorando los datos abiertos que hay en cada uno de los países. Obviamente, en los que nosotros vivimos, ¿no? También pueden estar investigando -- recuerdo el caso de colegas de Colombia que estaban investigando migración con sus fronteras, en este caso puede ser muy útil que en el caso de Colombia, los colegas colombianos busquen información de Venezuela en otros países, por ejemplo, si están buscando el tema de la migración venezolana. Así que los sitios de gobierno abierto son un buen recurso para comenzar con bases de datos.

[00:08:26] **Sandra Crucianelli** Después hay bases de datos que son de organismos internacionales. Una de las más conocidas es Finances. Finances es la base de datos de economía financiera del Banco Mundial, aquí hay datos abiertos sobre muchos países. Aquí en la zona de créditos y préstamos ustedes pueden ver cuáles son los más accesados, los que son de más frecuente consulta. Este, por ejemplo, que es el estado de los préstamos que ha asumido cada país. Y bueno, lo interesante de esto es que ustedes pueden descargar la información en distintos formatos, allí la tienen en formato de tabla, descargar, aquí tienen las opciones de exportación. Vean que cuando ustedes dentro de finanzas van al área de exportar, hay básicamente varios formatos de archivos y los más comunes son CSV y CSV for Excel. Si dice CSV for Excel, en general ya es un CSV -- yo les hablé de este tipo de archivo de atributos separados por comas, en este caso si lo descargan para Excel van a poder abrirlo con la aplicación y en algunos casos ya están estructurados, en otros casos van a tener que estructurarlo como en este por ejemplo, este es un CSV que se descarga fácilmente. Y recuerden que la forma de acceder un CSV es primero descargándolo, después convirtiéndolo en un Excel: marco la primera columna, voy a la pestaña datos, voy a clic texto en columnas, marco en la opción de delimitados, dejo marcadas las opciones de tabulación y coma, sigo al siguiente y le doy

clic en finalizar. Una vez que le di clic en Finalizar, aquí tienen un archivo bastante grande, tiene más de 300 filas y tiene muchas columnas. Como ustedes pueden ver, al final generalmente están los montos en la moneda de curso legal de acuerdo al tipo de país en el que ustedes estén viviendo. Toda esta información que está en Finances, obviamente está en dólares. Es una base de datos muy importante para todo periodista de datos.

[00:10:51] **Sandra Crucianelli** Pero también el Banco Interamericano de Desarrollo, por ejemplo, no solamente en el Banco Mundial, sino que el Banco Interamericano de Desarrollo tiene información en datos abiertos y vale la pena revisarla -- ustedes puedan buscar por país -- en este caso vamos a buscar Bolivia, hice clic sobre Bolivia. Y fíjense que hay muchos conjuntos de datos que ustedes se pueden descargar. También el Fondo Monetario Internacional tiene un área de datos descargables. Estos datos se pueden buscar también por país o por indicador, y todos son descargables. La mayoría descarga en Excel y en CSV. También la Organización Mundial de la Salud, supongo que quienes han estado cubriendo pandemia este año, bueno, yo la he usado mucho a esta base de datos. Ustedes pueden ir por tema, pueden ir por categoría, pueden ir por indicador, y lo interesante de esta base de datos es que también permite la descarga en formatos abiertos para todos los países que reportan a la OMS.

[00:12:06] **Sandra Crucianelli** Es habitual que los conjuntos de datos muy grandes, por ejemplo en la Universidad de John Hopkins --, que han puesto online con el comienzo de la pandemia su mapa de evolución de la pandemia --, tiene su repositorio en un sitio que se llama GitHub donde bueno, este sitio es para alojar conjuntos de datos muy grandes. Generalmente lo usan los programadores, pero ya en los últimos tiempos, en especial a partir de la pandemia, hemos aprendido a acceder a datos que están archivados y repositorios de esta naturaleza. Acá también van a encontrar mucha información en CSV. También la FAO, la Unión Europea, Naciones Unidas. Por ejemplo, Naciones Unidas tiene un conjunto de bases de datos -- la mayoría están en inglés y van a tener ciertas ventajas si leen este idioma con relación a si no lo leen porque hay muchas bases de datos internacionales con información de nuestro países -- Naciones Unidas, por ejemplo, también tiene un portal de datos abiertos sobre población y decenas y decenas de variables en estudio. Todos estos datos se pueden, fíjense que cuando yo selecciono en caso de Naciones Unidas, el formato de descarga me dice que puedo descargar en Excel, en este caso valores separados por coma, y también en otros formatos de archivo, pero siempre como que Excel y CSV están presentes.

[00:13:52] **Sandra Crucianelli** Estos son organismos internacionales, pero también tenemos en la pestaña de empresas un montón de recursos para encontrar información sobre compañías. Uno muy frecuente de uso es Open Corporates, que les permite buscar, vamos a suponer de Brasil, información de compañías -- en este caso hice la búsqueda solamente por un país, Brasil -- y estos datos me los puedo descargar en formatos abiertos también. En este caso, fíjense que las opciones de descarga no contemplan ni Excel ni CSV -- contemplan extensiones que son más bien de uso para los desarrolladores, programadores -- este archivo, este tipo de archivo que se llama JSON, es un archivo no conocido para muchos periodistas, pero ya vamos a ver cómo convertir un archivo JSON en un Excel. Ese va a ser uno de los desafíos del próximo módulo. Van a ver que es relativamente sencillo.

[00:14:59] **Sandra Crucianelli** También hay información sobre petroleras, compañías mineras, se puede en Resource Project que es una base de datos relativamente nueva, hace unos pocos años que está en línea. Lo bueno es que te permite seleccionar por

país, ¿sí? Yo puedo seleccionar un determinado país, voy a seguir con Brasil, y descargarme toda la información sobre los royalties, toda la información sobre impuestos, regalías y compensaciones que han hecho las mineras, básicamente al país en cuestión. Ahora está tardando un poquito en descargar, seguramente debido a la conexión, pero ahí la ven, ahí ven la ubicación de estas compañías, ven los stock que tienen esas compañías, y ustedes esta información se la pueden descargar en formato abierto tanto en CSV como en Excel a su computadora. Es una de las bases de datos sobre minería más completas que hay en todo el mundo. Y otras también, ¿no? Por ejemplo, DueDil es un recurso que tiene información sobre compañías y claro, desde los leaks del Consorcio Internacional de Periodistas de Investigación pueden buscar dentro de offshore leaks la base de datos de los leaks, de las offshores, pueden buscar por país. Y lo interesante que tienen -- yo voy a buscar en el caso de Argentina -- lo interesante que tiene esta base de datos es que ustedes pueden descargar este recurso en formatos abiertos. También se pueden llevar toda esta información a través de un CSV.

[00:16:46] **Sandra Crucianelli** Pero no solamente sobre compañías. Hay bases de datos, por ejemplo en España, los residentes en España con certificado de registro o tarjeta de residencia en vigor, si ustedes ven aquí, van a ver el iconito de Excel, es un Excel que te permite descargar el registro completo y bastante actualizado -- esto está a 31 de diciembre de 2020 -- es un recurso muy interesante especialmente cuando están investigando migraciones. También hay otros recursos, The Refugee Project, voy a ver The Migrant Files -- yo les seleccioné aquí base de datos que hemos estado usando en los últimos cuatro o cinco años y que tienen todas formatos abiertos de descargas. También la CEPAL tiene información sobre femicidios, un tema que ahora está ocupando muchos titulares en todos los medios de comunicación de todo el mundo. Este Observatorio de Igualdad de Género de América Latina y el Caribe tiene información en datos abiertos, posible de descarga.

[00:17:59] **Sandra Crucianelli** También hay aplicaciones de noticias que se han creado en los últimos años. Por ejemplo, este que tiene que ver con la compra de armas. Hay un sitio sobre armas, compras de equipamientos militares y temas de defensa, es el Sipri. Esta base de datos tiene aquí ustedes ven el Excel File, ustedes se pueden descargar aquí datos para todos los países del mundo, es una base de datos entre años 1949 y 2020. ¡Maravillosa esta base de datos! La descargan y sobre esos datos alguien ideó esto, que me encanta, ustedes podrían pararse sobre cualquier país -- en este caso lo estoy haciendo sobre Brasil -- para ver el comercio que hubo de armas. Puedo hacerlo sobre México, ¿sí? Y esta información que está contenida aquí, viene de aquí, ¿sí?, de esta base de datos. Por lo tanto, esto es una muestra de lo que podemos hacer, de las maravillas que podemos hacer con los datos abiertos, ¿sí? A partir de una base de datos que publica, con datos de Naciones Unidas, información sobre gastos en materia de defensa, básicamente son gastos en armas, el comercio en un color está lo que se exportó y en otro color lo que se importó. Y me parece que es una base de datos estupenda que les puede permitir hacer otro tipo de gráficos interactivos y llevar historias al medio de comunicación en la que ustedes estén trabajando. También hay bases de datos con información sobre crisis ambientales, en este caso el Sipri [SERVIR] para todos aquellos quienes viven en la región de Mesoamérica hay mapas en tiempo real y lo bueno que tienen estos sitios es que tienen datos abiertos descargables y ustedes podrían montar sus propios gráficos sobre ellos. El CRED también es otro recurso en materia de cobertura de crisis, el Centro de Investigación y de Desastres y Epidemiología de Desastres. Allí hay mucha información sobre terremotos, sobre inundaciones, sobre crisis que han tenido que ver con en distintas situaciones en distintas partes del mundo, y están los datos históricos que se los pueden descargar a su computadora.

[00:20:42] **Sandra Crucianelli** También información sobre tráfico marítimo. Hoy día hay plataformas Marine Traffic, por ejemplo, tiene sus sus datos abiertos, ustedes pueden ubicar aquí está moviendo en tiempo semireal, digamos, el delay es de uno o dos minutos, podemos hacer clic sobre uno de los buques, esto es sobre tráfico marítimo, y en algunos casos me da información del buque, un carguero, el Sea Beijing está transitando estas aguas jurisdiccionales. Incluso en algunos casos hay fotos del carguero y te muestra la trayectoria geográfica del punto del puerto de partida al puerto de llegada. Así como hay para los barcos --, podemos ubicar barcos en todas partes del mundo fíjense, el tráfico marítimo que hay en este momento es superior al tráfico aéreo --, tenemos el de vuelos: Flight Radar es la base de datos más usada. También tiene sitios..., tiene datos abiertos para descarga. Yo me voy a ir alejando, hay muy poco tráfico aéreo en mi país en estos momento, pero sí hay, voy a irme alejando, vean el mundo, ¿sí? Yo me puedo ir acercando a donde mayor tráfico aéreo hay que es en Estados Unidos y en Europa. Y cuando yo me posiciono sobre un avión en particular -- este es un avión de United que ha salido de Nueva York y se dirige a San Juan de Puerto Rico -- puedo ver la foto del avión, su ruta y el tiempo estimado de vuelo, y toda es información muy importante cuando estamos haciendo algunas investigaciones en las que necesitamos seguir ciertas trayectorias. Por lo tanto, en Vessel Finder, que es la otra que me quedaba por mostrarles, que también es de tráfico marítimo. Está Flight Radar, que es la que recién les compartí, para vuelos. Está Vessel Finder para buques y Marine Traffic también para buques.

[00:23:12] **Sandra Crucianelli** También hay bases de comercio exterior. Si ustedes entran al sitio web del censo de los Estados Unidos, Census.gov, van a poder encontrar todo lo que Estados Unidos exportó a distintos países. Yo, por ejemplo, voy a ver aquí Bolivia, y lo que van a tener es la progresión completa en miles de dólares de lo que Estados Unidos exportó a Bolivia en este caso. Y yo podría haber usado cualquier otro país, vamos a ver aquí Colombia, ¿ven? Esta información, si bien no está descargable, es muy fácil de llevártela a una hoja de cálculo, sólo tienen que marcar y con el botón derecho del mouse copiar y llevarme toda esta información a una hoja de cálculo. También hay formas de encontrar a personas en la web en algunos repositorios como Webmii que permite encontrar personas, y también Peek You en el caso de que ustedes necesiten ver de qué la gente está hablando en un determinado momento, es un recurso también es una base de datos interesante. Lo mismo que es Social Mention si ustedes están rastreando conversaciones en las distintas redes sociales, podrían buscar en Twitter o podrían buscar en Facebook, también en Instagram. Incluso Twitter, que es una de las redes sociales más usadas, tiene una búsqueda avanzada que es está: Twitter punto com barra, search, guion advanced, casi nadie lo usa, pero les puedo asegurar que esta plantilla de búsqueda es muy útil porque les permite recolectar conversaciones específicas de algunos usuarios o salir a la pesca de algunos temas que sean de su interés.

[00:25:17] **Sandra Crucianelli** A ver, en ese biblioato que yo les compartí, en ese LiveBinder de bases de datos, hay muchas bases de datos. No se preocupen por los links porque por ahí me preguntan "pero Sandra, mostraste este recurso y no sé dónde está la url". Bueno, en los textos de la clase de esta semana están todas las bases de datos que yo les mostré y muchas más están listadas. Ahora depende de ustedes. Ustedes tienen que accesarlas, ver qué es lo que hay allí, si encuentran información del país en el que viven y en el país en el que trabajan, es decir, ahora ya depende de ustedes. Esta es como una guía que pretende ayudarlos a encontrar datos en la gran autopista informática que tenemos en nuestros tiempos. Es todo un desafío. A mí me parece como ir a buscar

perlas al fondo del mar. Es una tarea muy interesante. Cuando uno está navegando, usando un explorador o un navegador, lo que uno está viendo es la superficie; cuando entran a las bases de datos se introducen en un nivel mucho más profundo de la web y como les digo, es buscar una perla en el fondo del mar. Así que en sus búsquedas, les deseo mucha suerte.