

Introducción	3
Datos abiertos	5
Introducción	5
¿Qué son los datos abiertos?	5
¿Para qué sirven los datos abiertos?	6
Breve historia de los datos abiertos	7
Estructura de los datos abiertos	8
Impacto de los datos abiertos	9
¿Cómo se relacionan los datos abiertos y gobierno abierto?	10
Datos abiertos y código abierto	15
Datos abiertos y big data	16
Resumen	17
Referencias	18
Datos abiertos en el mundo y en Argentina	20
Introducción	20
Escalas de la apertura de datos	20
Instituciones estandarizadoras	22
Métricas de los datos abiertos	24
Iniciativa de datos abiertos en el mundo	27
Iniciativa de datos abiertos en Latinoamérica y Argentina	32
Casos de éxito	35
Desafíos en la región	38
Resumen	41
Referencias	41
Datos abiertos en universidades	44
Introducción	44
¿En qué consisten?	44
¿Cuáles son sus beneficios?	45
Investigación abierta, acceso abierto y comunidad colaborativa	46
Iniciativa de datos abiertos universitarios en el mundo	48
Iniciativa de datos abiertos universitarios en Latinoamérica y Argentina	51
Casos de éxito	56
Resumen	57
Referencias	57
Desarrollo	59
Introducción	59
Herramientas utilizadas para el prototipo	59
Plataformas de publicación de datos abiertos	70
Análisis de beneficios del prototipo frente a otros portales de datos abiertos	74
Resumen	81
Referencias	82
Prototipo	84
Introducción	84

Conceptos generales del prototipo	85
Análisis de la apertura de datos	86
Informes de usuarios	95
Resumen	99
Referencias	100
Propuestas para futuros desarrollos	101
Anexo	101
Conclusiones	102
Bibliografía	104

Introducción

Hace ya varios años que el mundo se fue orientando hacia un paradigma que promueve la apertura de información para mejorar sus mecanismos de transparencia y aprendizaje, establecer nuevos puntos de interacción entre comunidades de diferentes temáticas, y fomentar la co-creación de nuevos e innovadores productos y servicios; este nuevo paradigma es el que pregonan las iniciativas de datos abiertos.

Se pueden definir a los datos abiertos, según la fundación Open Knowledge, como “... aquellos datos que pueden ser utilizados, reutilizados y redistribuidos libremente por cualquier persona”, pero quedarse con esta definición formal es limitarse en el conocimiento de un ecosistema aún más grande que comprende una filosofía de apertura total. La ideología de datos abiertos engloba no solo el resultado como información redistribuible o reutilizable, sino que contempla los procesos y mecanismos mediante los cuales fueron logrados esos resultados para que sean abiertos a una comunidad que aproveche esa información. Contempla, además, el marco normativo necesario para que los consumidores de los datos puedan llevar a cabo los principios antes mencionados sin restricciones, mediante licencias abiertas.

De las iniciativas de datos abiertos se desprenden varios conceptos interesantes, que desde hace ya un tiempo vienen ganando terreno en las diferentes áreas tanto públicas como privadas de los distintos países del mundo, tales como acceso abierto que intenta establecer la apertura de la información de revistas científicas las cuales se utilizan en varios programas de aprendizaje en universidades del mundo; y como investigación abierta, que fomenta la apertura de los procesos y elementos utilizados en el transcurso de la investigación hasta obtener el resultado. Pero el concepto más en mayor auge de las iniciativas de datos abiertos es el de gobierno abierto que comprende las políticas necesarias para que estas cumplan sus objetivos en cada una de las áreas gubernamentales. Gobierno abierto se encarga de fomentar la participación ciudadana; de inculcar los beneficios de apertura de la información, entre ellos la transparencia y la rendición de cuentas, para provecho no solo de los ciudadanos sino también para agilizar los procesos burocráticos en sus instituciones; de difundir y generar espacios de co-creación del estilo de hackatones, datatones e ideatones, con el fin de alcanzar una mayor adhesión de la comunidad tecnológica para fomentar la innovación de nuevos productos con los datos de las áreas de gobierno. Universidad abierta es también un concepto interesante, que engloba los tres anteriormente mencionados, focalizando su atención en la motivación de los estudiantes universitarios en toda su vida académica, mejorando las herramientas de aprendizaje con que cuentan los docentes de las instituciones y estableciendo una comunicación directa entre todos los actores involucrados del ámbito académico.

Uno de los medios de comunicación de datos abiertos es el de la utilización de una plataforma de publicación de datos abiertos, la cual debería permitir: la gestión de la información a disponibilizar; la asociación de visualizaciones a dichos datos que permitan facilitar un rápido pantallazo del contenido de la información al consumidor; la interacción de los consumidores de los datos con la información concreta; la catalogación de los diferentes conjuntos de datos provistos por la entidad involucrada; la disposición de secciones que establezcan una comunicación directa con el consumidor de los datos del estilo de comentarios, blogs o redes sociales; la descarga de la información directa a través de formatos abiertos o de servicios web vinculados; entre las características más relevantes.

En este documento se detallará el marco teórico sobre el concepto de datos abiertos y sus beneficios; cuál es el impacto de llevar a cabo una iniciativa de datos abiertos; su relación con los

conceptos de código abierto y big data y sus beneficios; una breve historia sobre sus comienzos y su vínculo directo con gobierno abierto; las instituciones estandarizadoras, las escalas de apertura y las métricas para determinar el posicionamiento de las diversas iniciativas de datos abiertos a nivel mundial; la incorporación de las iniciativas de datos abiertos en las universidades y sus beneficios; los casos de éxito utilizando datos abiertos correspondientes a los ámbitos gubernamentales y académicos. Por último se analizarán las diferentes herramientas de publicación de datos y se expondrán los motivos por los cuales se ha determinado la creación del prototipo comprendido en esta Tesina; se mencionarán las herramientas tecnológicas utilizadas y se mostrarán las comparativas confeccionadas entre las herramientas previamente analizadas; y se pondrán a disposición las evaluaciones realizadas por actores vinculados al área de datos abiertos sobre la plataforma desarrollada.

Datos abiertos

Introducción

Hablar de datos abiertos supone un nuevo paradigma, una nueva forma de conectarse al mundo y colaborar; incentiva a la idea de generar comunidades para compartir la información y obtener los recursos necesarios para la creación de nuevos e innovadores proyectos.

En este capítulo se abordarán los temas principales del concepto de datos abiertos y cómo estos se relacionan tanto con las políticas que llevan adelante los gobiernos, como también con otros conceptos que día a día toman mayor relevancia en el ecosistema informático como big data y código abierto.

Es acertado hacer hincapié en la evolución constante de este paradigma y en cómo impacta en el quehacer de la vida cotidiana de los ciudadanos cuando se abordan políticas de estado adecuadamente y consensuadas entre todas las áreas involucradas. No obstante, al ser un concepto novedoso, se presentan limitaciones en cuanto a la cantidad de áreas que llevan adelante este tipo de iniciativa, pero no hay dudas de que el resto no tardará en adoptar esta política de apertura al notar los beneficios que implica su aplicación.

¿Qué son los datos abiertos?

“Los datos abiertos son todos aquellos datos que pueden ser utilizados, reutilizados y redistribuidos libremente por cualquier persona.” En base a esta afirmación de la fundación Open Knowledge^[1], se puede entender a los datos abiertos como toda información disponibilizada por distintas entidades (instituciones, empresas, gobiernos, personas, etc.), para ser utilizada y reutilizada con distintos fines, como la generación de sistemas, documentos e investigaciones. A su vez, son redistribuidos de la misma forma en que fueron percibidos, es decir, completamente libres, sin ningún tipo de restricción, para el beneficio común.

Al hablar de datos abiertos, se debe hacer hincapié en tres características fundamentales:

- **Disponibilidad y acceso:** los datos deben estar disponibles en su conjunto y de forma detallada. La forma más conveniente de publicación es en internet con la opción de descarga libre y directa en diferentes formatos estándares como por ejemplo xls, csv o json.
- **Reutilización y redistribución:** los datos deben ser disponibilizados de forma en que puedan ser reutilizados y redistribuidos libremente, incluyendo la posibilidad de relacionarlos con otros conjuntos de datos.
- **Participación universal:** los datos pueden ser utilizados, reutilizados y redistribuidos libremente por cualquier entidad (personas, grupos, empresas, fundaciones, gobiernos, etc) sin discriminación por su campo de actividad.

En síntesis, todo dato abierto debe estar disponible de forma fácil y accesible. En un formato estandarizado y estructurado siguiendo los principios descritos por la Open Definition^[2].

Una de las grandes confusiones que surgen a la hora de hablar de datos abiertos es creer que todos los datos que se encuentran disponibles públicamente lo son, ya que existen datos publicados que no necesariamente se encuentran dentro de la categoría de datos abiertos. Sin embargo se puede afirmar que todos los datos abiertos se encuentran disponibles públicamente.

Es prácticamente imposible que un gobierno u otra entidad disponga públicamente de todos sus datos. Puede hablar de datos abiertos y de iniciativas de apertura de datos sin necesidad de hacer pública toda su información.

Existen regulaciones legales que rigen sobre los datos a la hora de su apertura. Por ejemplo, se debe asegurar que su licencia permita a los usuarios o consumidores su reutilización dando, o no, crédito a la fuente del dato. Además, existen restricciones legales en cuanto a la publicación de datos sensibles como los de identidad, ganancias privadas, facturación, etc.

Los avances tecnológicos de los últimos años son de gran importancia a la hora de generar herramientas para disponibilizar datos. Este progreso marca el camino a seguir en cuestiones de estandarización de formatos y estructuras, publicación y divulgación de los datos abiertos. Al contrario de la creencia popular, a la hora de abrir y publicar datos se debe tener en cuenta ciertas normas dentro de un estándar, por ejemplo la guía que propone Open Data Handbook^[3].

No todos los formatos son los adecuados para la apertura de los datos en cuestiones de datos abiertos, por ejemplo, en formato .pdf o .xls con imágenes, ya que dificultan la explotación de los mismos. La forma correcta de apertura es con la información clara y disponible para relacionar con otros conjuntos de datos. Además, el ideal, es que su divulgación sea mediante plataformas web que permitan acceso masivo al común de la gente.

Según el Data Revolution Group^[4], “Los datos son el alma del proceso de adopción de decisiones y la materia prima para la rendición de cuentas”. Ante esto, no cabe duda de que nos encontramos frente a una iniciativa ampliamente poderosa que brinda a la sociedad conocimiento de lo que sucede a su alrededor y le permite a las organizaciones tomar decisiones en base a evidencia.

¿Para qué sirven los datos abiertos?

Entre los beneficios que presentan los datos abiertos encontramos la performance de los servicios de los organismos que los utilicen. Permiten mejorar la eficiencia de los procesos y productos que llevan a cabo las diferentes entidades logrando por consiguiente una mayor aceptación, sobre dicha entidad, por parte de terceros. Asimismo, permiten la reducción de costos en los servicios aplicados, sobre todo en la conexión o comunicación entre los organismos para contar con la materia prima necesaria en la implementación de sus servicios, procesos o productos.

Los datos abiertos pueden convertirse en una herramienta para eliminar las brechas de información entre los organismos, permitiendo compartir puntos de inicio o referencia y difundir las mejores prácticas que elevan la productividad. Además, pueden impulsar la innovación en ejes transversales y ayudar a las entidades a replantearse los enfoques de toma de decisiones tradicionales basándose en datos. El análisis de datos abiertos también puede generar, o ayudar a descubrir, un conjunto de preferencias del consumidor, incentivando a las entidades a realizar la

productividad y descubrir anomalías o variaciones innecesarias de sus productos. Esto permite traducirse en procesos más sencillos y más confiables.

A su vez, permiten incluir el concepto de transparencia. Hablar de transparencia de los datos es un gran avance en las políticas internas y externas de las entidades que llevan adelante esta iniciativa, dando al público un pantallazo de lo que sucede en el día a día dentro de las organizaciones. La transparencia de los datos otorga un grado de visibilidad nunca antes alcanzado, por lo que se abre un paradigma que produce buenos frutos tanto para el productor como para el consumidor de los datos.

No se puede dejar de mencionar la explotación de los datos abiertos por parte de grupos académicos, gobiernos, empresas, pymes e investigadores de distinta índole. Estos grupos se ven completamente beneficiados al poder contar con datos útiles para sus investigaciones, toma de decisiones, conformación de nuevos mecanismos y desarrollos y la creación y/o integración de nuevos sistemas. Esto genera una línea de aprovechamiento de la información en cadena, donde cada uno de los eslabones que la conforman pondrá a disposición nuevos datos en forma abierta surgidos de sus proyectos, lo que supone una nueva etapa de mejora y avance.

Sin embargo, los beneficios planteados requieren inversiones en tecnología y experiencia del personal para que los datos sean utilizados de manera efectiva. Además, las organizaciones deben trabajar para diseñar y establecer políticas que protejan la privacidad y la propiedad intelectual, y generar estándares para acelerar el flujo de los datos.

Los beneficios de los datos abiertos pueden reforzarse a sí mismos ya que aumentarán a medida que las personas perciban las ventajas y ayuden a mejorar la precisión y el detalle de la información disponible. Sin embargo, este ciclo puede cobrar impulso solo si las organizaciones cultivan un ecosistema de datos abiertos e implementan las políticas antes descritas.

Breve historia de los datos abiertos

En primer término, se puede enunciar que el concepto de datos abiertos surge en el ámbito científico, más precisamente en los datos relacionados a los estudios geofísicos y ambientales en 1995. Dichos investigadores promovían la apertura de información y el intercambio masivo entre los diferentes países para entender los distintos fenómenos que surgían en el mundo. Los ideólogos de aquella época comenzaban a hacer hincapié en la liberación de los datos y el despojo de los derechos de autor con el fin de disponibilizar en un fondo común toda la información pertinente con el objetivo de lograr una continuidad en los proyectos iniciados en el ámbito científico. Cabe mencionar que estos inicios e ideologías tomaron real importancia cuando se logró unificar los conceptos con el avance tecnológico, específicamente con los ideales del software libre y el código abierto.

Tanto el término datos abiertos como la comunidad que se asocia a él, pueden ser considerados relativamente recientes. Apenas hace un poco más de diez años que dichos conceptos comenzaron a ganar popularidad producto de la determinación, por parte de un grupo de defensores del gobierno abierto, de establecer los ocho principios de los datos abiertos gubernamentales. Este suceso, que tuvo lugar en Sebastopol (California) en diciembre de 2007, perseguía el objetivo de definir los conceptos principales por los cuales diferenciar a los datos abiertos de los demás. Con este propósito establecido se lanzaron las iniciativas oficiales de datos abiertos por parte de los gobiernos de Estados Unidos (data.gov) y del Reino Unido (data.gov.uk), dándole el empujón final a una corriente que comenzaba a tener identidad propia.

El ex presidente de los Estados Unidos, Barack Obama, toma en consideración a este movimiento ciudadano, sus criterios y su escala, a fin de poner en marcha una política de liberalización de datos. En 2013, suscribe un decreto ejecutivo cuyas disposiciones tratan sobre la apertura de datos públicos bajo formatos que permiten la reutilización.

“La apertura en el gobierno fortalece la democracia, promueve la entrega de servicios eficientes y efectivos al público y contribuye al crecimiento económico (...) el valor predeterminado de los recursos de información gubernamental nuevos y modernizados deben ser abiertos y legibles por máquinas.”, expresó el ex presidente Barack Obama en 2013 en el *Decreto Ejecutivo 13642: Haciendo el nuevo valor predeterminado de la información gubernamental de información gubernamental abierta y legible por máquinas*^[5].

Estructura de los datos abiertos

Es esencial entender que los datos abiertos se deben distribuir en formatos estándares y abiertos con el fin de lograr unificar la estructura y facilitar así su conceptualización, su análisis y su posterior explotación. Inicialmente se debe interpretar por formato a la estructura en la que un conjunto de datos se almacena. Por otra parte, el concepto de apertura de dicho formato hace hincapié en disponibilizar la documentación de su estructura e implementación.

Estos conceptos dan lugar a mejoras continuas sobre las herramientas con las que se puede disponibilizar a los datos abiertos. Permiten mejorar la compatibilidad entre las diferentes plataformas y el uso irrestricto de las mismas.

Los formatos abiertos no son los únicos que se utilizan al momento de disponibilizar los datos, también suele hacerse en formatos cerrados o propietarios. Las restricciones de utilizar estos formatos conlleva la dependencia con los softwares de terceros o con los propietarios de las licencias que los regulan. En la sección anterior se definieron los puntos más importantes que deben contemplar los datos para ser abiertos, por lo que el uso de estos formatos para disponibilizar los datos implicaría una contradicción teórica.

Algunos de los formatos más utilizados para la apertura de la información son:

- **Datos tabulares:**
 - **CSV (TSV):** Archivo de texto plano para representar información tabulada. Cada columna se separa mediante una coma (o tabulación). No es un formato estandarizado ni tiene una estructura predefinida, por lo que a veces es necesario un documento que explique la estructura del dato.

- **Datos estructurados:**
 - **XML:** Lenguaje de marcado que define un conjunto de reglas con lo que se permite mantener un estándar de estructura y definir un estándar para la creación.
 - **JSON:** Lenguaje de notación de objetos muy simple de leer por cualquier lenguaje de programación. Se convirtió en un estándar para compartir información entre sistemas.

- **XLS:** Formato propietario. Hoja de cálculo que muestra información en celdas organizadas en filas y columnas.
- **RSS:** Formato XML utilizado para distribuir contenido y actualización de ese contenido en la web.
- **Documentos de texto:**
 - **TXT:** Documento de texto sin ningún tipo de estructura que generalmente se utiliza para incluir metadatos.
 - **HTML:** Lenguaje de marcado para la creación de páginas web. Se utiliza para datos estables y de alcance limitado ya que no es muy amigable para compartir.
 - **Otros:** Documentos en formato Word, PDF, ODF, de los cuales únicamente el último es un formato abierto, que permiten crear hojas de cálculo, textos, gráficos y presentaciones.
- **Datos espaciales:**
 - **KML:** Archivo XML que permite mostrar información geográfica.
 - **SHP:** Formato propietario de datos espaciales que se convirtió en un estándar de facto^[6].

Impacto de los datos abiertos

Los datos abiertos cuentan con una amplia fuente de origen pero los dos mayores campos de aplicación y de publicación de datos son la ciencia y el gobierno.

A continuación se mencionan algunos de los proyectos con mayor impacto social, sin profundizar mucho en los casos relacionados a gobierno abierto, los cuales se abordarán en el siguiente capítulo con más detalle.

El ejemplo de mayor preponderancia de datos abiertos aplicado a la ciencia, fue el Proyecto del Genoma Humano^[7] cuyo resultado determinó la secuencia de pares de bases químicas que componen el ADN y logró identificar los 20.000-25.000 genes del genoma humano desde un punto de vista físico y funcional. Esto permitió entender enfermedades de mutación genética y distintos tipos de cáncer, entre otras; y en base a eso, se comenzaron a fabricar medicamentos más eficientes para estas afecciones. Al finalizar ésta investigación, todas las secuencias de ADN se publicaron y se encuentran actualmente disponibles para cualquier persona que las quiera consultar.

Otro caso interesante para ejemplificar fue el que surgió ante el brote de Ébola en Sierra Leona^[8]. Tres iniciativas: Sierra Leone's National Ebola Response Centre (NERC), Humanitarian Data Exchange (HDX)^[9] y Ebola GeoNode, cruzaron sus datos para llevar adelante diferentes proyectos para hacerle frente a la enfermedad. El de mayor impacto fue un mapa humanitario^[10] donde fácilmente se podían visualizar cuáles eran las áreas más afectadas por esta enfermedad, lo que permitió tomar decisiones para enfrentar la crisis sanitaria y evitar una mayor propagación del Ébola.

Cabe destacar el proyecto^[11] realizado en Canadá utilizando datos abiertos. En este caso, entidades de caridad hicieron públicos sus datos y al ser cruzados con otros conjuntos de datos permitieron explorar el estado del área sin fines de lucro y entender la relación entre los reguladores y las organizaciones benéficas para fomentar el apoyo y generar productos inteligentes para donaciones y recaudación de fondos.

Un caso muy particular surgió luego del sismo ocurrido en México en el año 2017. Un grupo de personas, pertenecientes a la organización Codeando México^[12], creó un sitio^[13] para concentrar los diferentes medios de apoyo tras el siniestro y proveer información rápidamente a quienes pudieran necesitarla; pero este proyecto que inició como un sitio web, terminó abarcando mucho más. Se realizaron mapas colaborativos, bots que identificaban noticias falsas, una aplicación para reporte de personas y animales perdidos y sitios de gestión de casos y asesoría jurídica, entre otros. Entre los valores más importantes de la realización de este proyecto se encuentran la participación ciudadana con fin solidario y el hecho de que estos sistemas tienen código abierto y están publicados en la plataforma Github, con lo cual cualquier persona puede sumarse al proyecto y colaborar. En un futuro, la idea de este proyecto, podría ser utilizada como base para catástrofes naturales de cualquier tipo.

También se pueden encontrar casos en el sector privado. Por ejemplo, algunas empresas como Enedis^[14], fomentan su transparencia publicando sus datos financieros en un portal y generando en base a dichos datos una web interactiva.

Si bien en el próximo capítulo se profundizará en el tema, cabe mencionar el gobierno abierto como medio para fomentar la transparencia de los estados. Entre los proyectos comunes podemos destacar los portales de datos abiertos de los distintos gobiernos a lo largo del mundo u Open Contracting Partnership^[15], quien impulsa la publicación de contrataciones realizadas por los estados.

¿Cómo se relacionan los datos abiertos y gobierno abierto?

La noción del término datos abiertos y específicamente de datos abiertos de gobierno - toda información, sea pública o no, que es de libre acceso y reutilizable para diferentes propósitos - viene creciendo desde hace ya varios años. Se debe hacer foco sobre la visibilidad de las políticas de datos abiertos en gobierno a partir del año 2009 con la aparición del Memorandum^[16] aplicado por el ex presidente de los Estados Unidos, Barack Obama, y el anuncio de las iniciativas hacia la apertura de gobiernos como Gran Bretaña, Canadá y Nueva Zelanda.

En primer término, se debe hacer hincapié en que datos abiertos no es lo mismo que datos públicos^[17]. Mientras que los datos públicos son los datos generados en el ámbito gubernamental, o que se encuentran bajo alguna de sus concepciones, y a su vez presentan diferentes restricciones para su divulgación como es el caso de la Ley de Protección de datos personales^[18], los datos abiertos son todos aquellos datos, públicos o privados, que se disponibilizan a través de un entorno digital, bajo una licencia abierta y con un formato abierto, y que no tienen el acceso restringido bajo ningún aspecto legal.

Se debe considerar la aplicación de la política de los datos abiertos como una herramienta sumamente productiva para quien la utilice. Tal es el caso de los gobiernos que, día a día hacen mayor hincapié en participar de las normativas que rigen sobre la apertura de los datos para fomentar y asentar lo que se conoce como Gobierno Abierto.

Los datos abiertos de gobierno se rigen por ocho principios^[19] definidos en consenso por la OGP (Open Government Partnership)^[20]:

1. **Completos:** Todos los datos públicos se ponen a disposición, sin limitaciones de privacidad, seguridad o privilegios válidos.
2. **Primarios:** Los datos se obtienen de la fuente, con el más alto nivel de granularidad, sin agregaciones ni modificaciones.
3. **Oportunos:** Los datos se pondrán a disposición tan rápido como sea necesario para preservar su valor.
4. **Accesibles:** Los datos deberán estar disponibles para la mayor cantidad de usuarios con la amplia variedad de propósitos.
5. **Procesables por máquinas:** Los datos deberán estructurarse razonablemente para permitir el procesamiento automatizado.
6. **No discriminatorios:** Los datos deberán estar disponibles para cualquier tercero, sin necesidad de registro previo.
7. **No propietarios:** Los datos deberán estar disponibles en un formato sobre el cual ninguna entidad tenga el control exclusivo.
8. **Libres de licencia:** Los datos no deben estar sujetos a ningún derecho de autor, patentes, marcas o regulación. Solo se permiten restricciones razonables de privacidad y seguridad.

Es importante entender que las políticas establecidas a nivel general por cada uno de los países, promueven la generación de portales de datos abiertos que cuentan con las tecnologías necesarias y estandarizadas para alcanzar una mayor eficacia en las publicaciones de información. Cada gobierno, ya sea nacional o local, provee una plataforma donde se visualizan los conjuntos de datos, normalizados y analizados previamente por las áreas específicas correspondientes, y establecen las estructuras en las cuales la información será presentada. Este proceso demanda una gran cantidad de pasos a seguir por parte de los productores de los datos, que consisten en:

- **Identificar los datos:** se debe hacer un análisis exhaustivo de los datos que se quieren disponibilizar, teniendo en cuenta cuáles son las áreas que los concentran, cuáles son las posibles visualizaciones que el dato permite realizar y poder normalizarlos.
- **Seleccionar los datos:** en este caso se requiere un análisis específico sobre criterios de valor (para actores internos y externos), costos de publicación y mantenimiento y riesgos asociados a la publicación. Estos criterios permiten evaluar el impacto que tiene la apertura de los datos no solo a quien los disponibiliza, sino también al consumidor de esos datos. Cuando se habla de valor de los datos, se debe considerar cuáles son los actores involucrados, cuántos de estos se ven beneficiados con la apertura, cuáles son las limitaciones de los actores sobre el uso de los datos; además, se debe hacer hincapié en evaluar el potencial uso de esa información, si permite una mejora en la eficiencia de los servicios, si genera una reducción de costos, si fomenta o establece oportunidades de empleo. Con respecto al concepto de costos, se deben estimar tiempos de análisis y preparación de los datos para su apertura, verificar que las tecnologías sean las óptimas para la disponibilización e identificar qué organizaciones comprometerán recursos humanos para el mantenimiento de la información. Por último, el riesgo de los datos pone en juego la confiabilidad de la institución por posibles interpretaciones erróneas; debido a esto se deben evaluar los aspectos legales de la apertura y su impacto sobre individuos o grupos que se puedan encontrar vulnerables ante dicha iniciativa.

- **Aplicar una licencia abierta:** como bien se mencionó, los datos deben disponibilizarse bajo condiciones legales abiertas, por lo que se debe escoger una licencia abierta que regule la apertura. Para ello existe una variedad de licencias que se encuentran disponibles para utilizar:
 - **Open Data Commons Public Domain Dedication and License o PDDL^[21]**, que permite difundir, reutilizar o adaptar los datos sin restricción alguna.
 - **Open Data Commons Attribution License^[22]**, exige la referencia a la autoría o fuente de los datos para la reutilización de la información.
 - **Open Data Commons Open Database License (ODbL)^[23]**, que permite la reutilización de los datos siempre que se reconozca la autoría de la información original y se mantenga la misma licencia en las obras derivadas, las cuales puede restringir su uso si, además, se distribuye una versión sin dichas restricciones.
 - **Creative Commons Atribución 4.0 Internacional^[24]**, que posibilita no sólo copiar y redistribuir el material en cualquier medio o formato, sino también adaptar y transformarlo para cualquier propósito, incluso comercial.
- **Abrir los datos:** luego de seleccionar exhaustivamente los datos y la licencia con la que se van a disponibilizar, se debe establecer el catálogo de datos, que es la herramienta principal para facilitar la apertura, presentarla de manera centralizada y permitir la realización de búsquedas sobre los diferentes conjuntos de datos. Hoy en día, se encuentran varios desarrollos de software que permiten esta funcionalidad, como CKAN^[25] (utilizado por el gobierno de Reino Unido, <https://data.gov.uk/>), Socrata^[26] (utilizado por la ciudad de Chicago, <https://data.cityofchicago.org/>), Junar^[27] (utilizado por el municipio de Bahía Blanca, <http://datos.bahiaablanca.gob.ar/home>) o DKAN^[28] (utilizado por la ciudad de California, <https://data.ca.gov/>). Todas estas herramientas serán detalladas en futuras secciones.
- **Promover el uso:** es fundamental que los datos sean alcanzados por los consumidores con facilidad y que existan comunidades para que la reutilización de dicha información se transforme en la generación de futuros productos. La difusión de jornadas de participación ciudadana como hackatones, ideatones y datatones son el claro ejemplo de estos objetivos que se intentan alcanzar, y obtener en consecuencia, prototipos e investigaciones que alimenten el ecosistema de datos abiertos. Por este motivo se debe fomentar la comunicación y visualización del esfuerzo alcanzado por las áreas competentes con el fin de lograr estos objetivos y contactar con las organizaciones que mayor uso prestarán sobre los datos, como desarrolladores, visualizadores, economistas, periodistas, estadísticos, entre otros.

Estos últimos cuatro procedimientos fueron propuestos por la fundación Open Knowledge International y conforman un estándar en el proceso de apertura de la información.

Cabe mencionar que uno de los aspectos importantes de las políticas de gobierno abierto es la rendición de cuentas y la transparencia. A través de esta, los gobiernos se ponen a disposición del control constante, por parte de los ciudadanos y/o de las entidades reguladoras, sobre los gastos financieros, los hitos prometidos, las contrataciones realizadas y el resto de sus actividades.

La iniciativa de gobierno abierto se desprende de la idea de garantizar al ciudadano la disponibilidad de todos los datos con los que cuenta cada gobierno para su posterior utilización, ya

sea un desarrollo informático, planillas de información o simplemente el conocimiento tácito de lo que ocurre detrás de la toma de decisiones que impactan en su vida cotidiana

El primer aval de esta política es la transparencia. Llevar a cabo la iniciativa de datos abiertos dentro del ámbito gubernamental fomenta un gobierno transparente en el que se encuentra disponible todo tipo de información, incluyendo los datos sensibles como pueden ser los de carácter financiero o económico, entre otros.

Al hablar de transparencia dentro de Gobierno Abierto se debe hacer mención a la diferencia que existe entre transparencia activa y transparencia pasiva^[29]. Se entiende por transparencia activa a aquella que constituye al Estado a tomar una iniciativa propia en lo que respecta a la disposición de los datos públicos sin que hayan sido solicitados previamente. Estos suelen ser publicados en sitios web específicos que facilitan el acceso, búsqueda, redistribución y reutilización de los mismos, tales como los portales de datos abiertos. Por otra parte, la transparencia pasiva involucra principalmente el establecimiento de un canal institucional para que los ciudadanos puedan buscar, solicitar y acceder a la información que produce el Estado.

Gobierno Abierto implica una política de apertura hacia el ciudadano a través de la publicación de información, la adopción de procesos de participación y el desafío a los ciudadanos para que se involucren en la búsqueda de soluciones innovadoras a problemáticas públicas. Esta iniciativa no solo beneficia la transparencia de un gobierno, sino que fomenta la participación ciudadana. El gobierno disponibiliza plataformas y medios de comunicación con los cuales los ciudadanos interactúan con ellos, aportando ideas o sugerencias, a partir de los cuales se pueden mejorar los servicios públicos.

Un gobierno colaborativo es el que abre el juego y propone trabajar en conjunto con la ciudadanía, las organizaciones no gubernamentales, emprendedores, universidades y empresas para encontrar soluciones innovadoras a las problemáticas públicas. La creación de planes y políticas para mejorar la calidad de vida de los ciudadanos llega a mejores resultados al realizarse junto a expertos de diferentes áreas y los propios usuarios. De esta forma, la solución a la que se llegue será en base a la experiencia y conocimientos de equipos multidisciplinares.



Relación entre los tres beneficios principales de la apertura de datos.

En el diagrama realizado por el equipo de Gobierno Abierto de la Provincia de Buenos Aires^[29], se ve con claridad la conexión entre los tres beneficios principales de la apertura de datos de gobierno y cuáles son los beneficios alternativos de cada uno de ellos.

Ponderar la política de Gobierno Abierto trae como consecuencia fortalecer las políticas públicas. La apertura de datos permite encontrar información complementaria en otras áreas. Mientras más organismos trabajen con catálogos estandarizados más rica será la información disponible y mejores las políticas públicas. A su vez, impulsa la innovación y el desarrollo económico.

La disponibilidad de datos genera oportunidades de empleo, ya que fomenta a las instituciones a generar áreas especializadas en gobierno abierto, contar con personal capacitado para el análisis, la normalización y disponibilización de los datos y áreas específicas para la difusión de las políticas adoptadas sobre los mismos. A su vez, permite la creación de equipos de auditoría para regular las normas que rigen sobre la iniciativa de datos abiertos. Además, emprendedores y ONGs pueden usar la información disponible para desarrollar nuevos servicios o generar nuevos conocimientos que permitan mejorar la calidad de vida de los ciudadanos.

Los datos en un gobierno abierto promueven cambios significativos en la toma de decisiones y las cuestiones operativas de la gobernabilidad, por ende tienen un efecto positivo en la vida de los ciudadanos. Esta iniciativa logra un gobierno más eficiente donde no hay duplicidad de datos, gastos innecesarios para la obtención de los mismos ni interpretación unívoca de la información; y se proporciona información de forma más precisa y permite alertar sobre datos faltantes.

Por otro lado, dentro de un gobierno, al generar gran cantidad de datos de carácter sumamente heterogéneo, surge el desafío de intentar relacionarlos con el fin de analizar y prever situaciones. Por ello es que dichos gobiernos impulsan encuentros en los que participan personas ligadas al ambiente IT, programadores, analistas o expertos en el desarrollo de aplicaciones, con el fin de afianzar y prosperar en la innovación de nuevos portales ciudadanos, que los acerque cada vez más y de forma más simple al ciudadano que no necesariamente corresponda al área de datos.

Datos abiertos y código abierto

Cuando hablamos de datos abiertos y código abierto, o cuando se menciona datos abiertos y software libre, se deben aclarar algunas confusiones que suelen presentarse ante la supuesta similitud de los términos que se emplean.

Usualmente se habla de datos abiertos y código abierto y de datos abiertos y software libre indistintamente, sin conocer que entre código abierto y software libre existen al menos dos diferencias principales. Una de ellas es de aspecto teórico, por un lado software libre hace referencia a cuestiones filosóficas, éticas y morales; mientras que el movimiento código abierto prioriza el aspecto técnico. La otra corresponde al ámbito comercial, indica que se puede obtener remuneración de una aplicación que fue desarrollada bajo estándares del software libre, ya sea por desarrollo, mantenimiento o soporte, siempre y cuando se entregue el código fuente. En cambio en los desarrollos realizados en código abierto no existe la obligación de la entrega del código fuente. Continuando con ésta línea, todos los productos, sus modificaciones y sus derivados desarrollados con software libre, tienen que mantener éste licenciamiento, mientras que el concepto de código abierto no obliga a realizarlo.

Si bien las diferencias existen entre estos términos, juntos permitieron el desarrollo de software de calidad gracias a las libertades que otorgan y a la colaboración de la comunidad.

“El software libre es el que respeta la libertad y la comunidad de usuarios. Es un asunto ético, no de comodidades prácticas”^[30], sostuvo Richard Stallman^[31] fundador del Movimiento por el Software Libre.

Un programa es software libre cuando los usuarios tienen las cuatro libertades esenciales^[32]:

- La libertad de ejecutar el programa como se desee, con cualquier propósito (libertad 0).
- La libertad de estudiar cómo funciona el programa y cambiarlo para que haga lo que quiera (libertad 1). El acceso al código fuente es una condición necesaria para ello.
- La libertad de redistribuir copias para ayudar a otros (libertad 2).
- La libertad de distribuir copias de sus versiones modificadas a terceros (libertad 3). Esto le permite ofrecer a toda la comunidad la oportunidad de beneficiarse de las modificaciones. En este caso, nuevamente el acceso al código fuente es una condición necesaria para ello.

La principal ventaja de la utilización de software de código abierto está dada por la capacidad de generar un escenario de colaboración con el proveedor de software o el integrador, poniendo foco en el desarrollo de una sociedad centrada en el negocio.

Muchos de los proyectos de datos abiertos prefieren la filosofía de software libre para garantizar el mejoramiento de las herramientas y la continuidad del proyecto. Así también hay muchas empresas tecnológicas que están optando por estas modalidades.

Juan Bauer, uno de los impulsores de la comunidad de código abierto en Paraguay, dijo: "Si consideramos que el open source nos permitirá modificar y mejorar la herramienta con la que procesamos nuestros datos, garantizando de esta manera a su vez que la misma no dependerá de una empresa o institución para su continuidad a futuro, creo que la mejor apuesta para los datos abiertos es utilizar herramientas open source o software libre. Por otro lado por una cuestión filosófica los datos abiertos deberían estar estrechamente ligados al open source y el software libre. Haciendo una analogía de los datos abiertos con el software privativo me imagino liberando la receta de una comida pero donde uno de los ingredientes solo es proveído por una empresa privada."^[33]

“El hecho de ser abierto también permite adaptar el software a necesidades específicas en casos particulares[...]. Las herramientas de software con código desarrollado y revisado por grandes comunidades de usuarios, también entrega un nivel alto de confianza, pues este código ha sido revisado y testeado por muchos más usuarios que los de una empresa particular.[...]”^[34], expresó Rainer Gross, gerente de productos de Sonda.

El desarrollo de herramientas de código abierto no sólo permite ganar escala y agilidad en la labor tecnológica, sino que también facilita el fortalecimiento de la comunidad de datos dentro y fuera de gobierno.

Por dichos motivos, los desarrollos pertinentes a la hora de disponibilizar los datos abiertos se focalizan en utilizar software de código abierto, no solo por la colaboratividad que dicho desarrollo de software presenta - su bajo costo y su mejora continua - sino también por evitar esa contradicción que presentan ambos términos. Se puede entender que tanto datos abiertos como código abierto van de la mano o así deberían hacerlo.

Datos abiertos y big data

Si bien los conceptos de datos abiertos y big data^[35] están estrechamente relacionados, hay que hacer especial énfasis en la diferencia que los ampara. Mientras que los datos abiertos se distinguen por su uso, big data se define por su volumen. El concepto de datos abiertos ya fue detallado, por lo que se hará foco en big data, que se define como grandes conjuntos de datos que no pueden tratarse de manera convencional ya que superan las capacidades de las herramientas utilizadas comúnmente para su análisis y procesamiento. Este concepto aplica no sólo al volumen de información con que se cuenta, sino también a la gran variedad de datos que pueden ser representados, ya sean a través de sensores, sistemas gps, dispositivos móviles, automóviles o medidores eléctricos, cuya velocidad de las aplicaciones que se encarguen del análisis de toda esta información deberá ser lo suficientemente rápida para tomar las decisiones más efectivas en el momento preciso. No sería adecuado quedarse únicamente con la asociación inmediata entre big data y “gran volumen de datos”, big data es mucho más que eso, trata sobre el valor que se puede obtener de los datos adquiridos, sobre el significado que estos tienen en un contexto determinado. Además, se debe contar con un ecosistema adecuado para este paradigma que surge de la coordinación entre la adquisición de la información, su almacenamiento y la tecnología con la que sea hará el posterior análisis y explotación de la misma.

En los últimos años se ha generado un crecimiento exponencial de los datos producidos y de los datos compartidos. Como se mencionó previamente, la ciencia y los gobiernos son las ramas que más aporte hacen tanto a los datos abiertos como al big data. Si sumamos este crecimiento con la reducción en el coste de procesamiento y almacenamiento, es posible tener información detallada sobre casi cualquier aspecto. El ejemplo más claro de este aumento en la producción de datos es la instalación de sensores en las ciudades inteligentes, los cuales generan datos en tiempo real sobre diferentes variables. La explotación de esta información permite mejorar, por ejemplo, la calidad del aire, el servicio de emergencias o el transporte, entre otros.

Uno de los principales pilares del gobierno abierto es fomentar la participación ciudadana desde la co-creación. Como se mencionaba anteriormente, existe un incremento de datos disponible para su explotación, lo cual permite a los gobiernos aprovecharlos y tomar decisiones basándose en

evidencia, es decir, en lo que realmente está pasando. Estos proyectos, adaptados a la necesidad de los usuarios, permiten tomar decisiones más inteligentes y acertadas.



Este gráfico muestra la estrecha relación que existe entre big data, datos abiertos y gobierno abierto, y qué datos se exponen de cada uno de ellos.

Si se logra explotar el poder de análisis que tiene el big data y la libertad que nos brindan los datos abiertos, se pueden transformar los negocios, gobiernos y la sociedad. Mientras que con big data podemos entender y analizar el mundo en el que vivimos; los datos abiertos nos aseguran que esto sea compartido y libre para quien quiera consultarlo.

Resumen

En consecuencia, este nuevo paradigma plantea una mejora significativa en la comunicación y co-gestión de proyectos independientemente del área al que corresponda; establece un enfoque colaborativo sobre los productos y servicios de las diversas temáticas vinculadas al procesamiento de la información; fomenta la participación ciudadana y la búsqueda constante de mejoras sustanciales en los servicios prestados por parte de los gobiernos.

Por otra parte, establece vínculos concretos con los conceptos de software de código abierto con el fin de ampliar la comunidad tecnológica y las áreas abocadas al análisis y gestión de los datos. Permite aprovechar la marcada tendencia de big data mediante herramientas de explotación y transformación de datos, lo cual tiene como consecuencia la focalización en la generación de información útil para el área que la lleve a cabo. Además, permite la innovación y la organización de eventos tecnológicos a disposición de los ciudadanos. Sin dudas, las iniciativas de datos abiertos se

irán expandiendo en las áreas que aún no las implementan a medida que las entidades responsables conozcan el impacto de sus beneficios.

Referencias

- [1] <https://okfn.org/>, fecha de consulta: 5 de diciembre 2017.
- [2] <http://opendefinition.org/>, fecha de consulta: 5 de diciembre 2017.
- [3] <http://opendatahandbook.org/>, fecha de consulta: 5 de diciembre 2017.
- [4] <http://www.undatarevolution.org/>, fecha de consulta: 5 de diciembre 2017.
- [5] <https://obamawhitehouse.archives.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government->, fecha de consulta: 12 de diciembre 2017.
- [6] https://es.wikipedia.org/wiki/Est%C3%A1ndar_de_facto, fecha de consulta: 27 de abril 2018.
- [7] <https://www.genome.gov/27562859/panorama-general-del-proyecto-del-genoma-humano/>, fecha de consulta: 12 de febrero 2018.
- [8] <http://odimpact.org/case-battling-ebola-in-sierra-leone.html>, fecha de consulta: 15 de febrero 2018.
- [9] <https://data.humdata.org/>, fecha de consulta: 15 de febrero 2018.
- [10] <https://www.gfdr.org/ebola-geonode>, fecha de consulta: 15 de febrero 2018.
- [11] <http://odimpact.org/case-opening-canadas-t3010-charity-information-return-data.html>, fecha de consulta: 16 de febrero 2018.
- [12] <https://www.codeandomexico.org>, fecha de consulta: 18 de febrero 2018.
- [13] <http://sismomexico.org/>, fecha de consulta: 18 de febrero 2018.
- [14] <http://www.enedis.fr/le-bilan-electrique>, fecha de consulta: 18 de febrero de 2018.
- [15] <https://www.open-contracting.org/>, fecha de consulta: 18 de febrero de 2018.
- [16] <https://www.archives.gov/files/cui/documents/2009-WH-memo-on-transparency-and-open-government.pdf>, fecha de consulta: 17 de diciembre 2017.
- [17] https://www.argentina.gob.ar/sites/default/files/2._kit_de_datos_abiertos.pdf, fecha de consulta: 18 de diciembre 2017.
- [18] <http://test.e-legis-ar.msal.gov.ar/leisref/public/showAct.php?id=14402>, fecha de consulta: 26 de junio 2018.
- [19] https://public.resource.org/8_principles.html, fecha de consulta: 19 de diciembre 2017.
- [20] <https://www.opengovpartnership.org>, fecha de consulta: 19 de diciembre 2017.
- [21] <https://opendatacommons.org/licenses/pddl/>, fecha de consulta: 26 de marzo 2018.
- [22] <https://opendatacommons.org/licenses/by/>, fecha de consulta: 26 de marzo 2018.
- [23] <https://opendatacommons.org/guide/>, fecha de consulta: 26 de marzo 2018.
- [24] <https://creativecommons.org/licenses/by/4.0/deed.es>, fecha de consulta: 26 de marzo 2018.
- [25] <https://ckan.org/>, fecha de consulta: 10 de marzo 2018.
- [26] <https://socrata.com/>, fecha de consulta: 10 de marzo 2018.
- [27] <http://www.junar.com/>, fecha de consulta: 10 de marzo 2018.
- [28] <https://getdkan.org/>, fecha de consulta: 10 de marzo 2018.
- [29] <https://www.gba.gob.ar/file/Kit%20de%20Apertura%20Municipal%202017.pdf>, fecha de consulta: 21 de diciembre 2017.
- [30] https://www.dcc.uchile.cl/richard_stallman, fecha de consulta: 16 de febrero 2018.
- [31] https://en.wikipedia.org/wiki/Richard_Stallman, fecha de consulta: 16 de febrero 2018.
- [32] <https://www.gnu.org/philosophy/free-sw.es.html>, fecha de consulta: 16 de febrero 2018.
- [33] <http://www.datos.org.py/open-source-software-libre>, fecha de consulta: 12 de marzo 2018.

[34]

<http://searchdatacenter.techtarget.com/es/cronica/Codigo-abierto-beneficios-novedades-y-consejos-de-implementacion>, fecha de consulta: 12 de marzo 2018.

[35] <https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/>, fecha de consulta: 29 de abril 2018.

Datos abiertos en el mundo y en Argentina

Introducción

A lo largo del tiempo la utilización de datos abiertos generó un amplio y variado campo de aplicación que día a día va incrementando y mejorando la colaboración entre quienes lo utilizan. Para lograr este paulatino avance se necesitó la participación, resolución y consenso entre diferentes instituciones y organizaciones que continuamente dan visibilidad a metodologías y estándares en el uso de la información abierta. A su vez, fomentan la participación constante de las áreas involucradas correspondientes a llevar adelante las políticas necesarias sobre este paradigma, con el fin de regular y concientizar sobre los beneficios e impactos que su implementación permite.

En este capítulo no solo se verá en detalle cada una de estas entidades estandarizadoras y las métricas que rigen sobre los datos abiertos, sino que se hará foco también en cómo la utilización de los mismos, con el correr de los años, ha incursionado en las políticas internas de los gobiernos y ha impactado directamente en la vida cotidiana de los ciudadanos. Se mencionarán los comienzos de la utilización de datos abiertos en el mundo y latinoamérica, y cuáles son los desarrollos y proyectos de mayor éxito utilizando este paradigma. Además, se listarán los portales de datos abiertos más relevantes junto con un breve análisis. Por último, se detallarán los desafíos propuestos para un futuro próximo por cada uno de los gobiernos de latinoamérica para mejorar sus políticas de datos abiertos y progresar en la transparencia de sus instituciones.

Escalas de la apertura de datos

Cabe destacar las diferentes escalas y parámetros que definen el nivel o grado de apertura de los datos. Cada nivel o escalón de desarrollo representa una mejora en los aspectos que conciernen a la disponibilización de los datos, ya sea el formato, la arquitectura empleada o las políticas respectivas de apertura. Este sistema busca diagnosticar el nivel de apertura de datos de los organismos públicos y provee grados alcanzables cada vez más refinados.

Estas escalas más conocidas como las *5 Estrellas de apertura de datos*^[1] (Five Start of Openness en su nombre original), fueron propuestas por Tim Berners-Lee^[2], el hombre más preponderante de Internet, director de W3C^[3] y fundador de WWW Foundation^[4], entre otros; y tienen las siguientes cualidades:

- **Primera estrella:** Corresponde a la disponibilidad de los datos en internet sin restricciones sobre el formato utilizado mientras cuente con una licencia abierta que permita la libertad de uso de la información.
- **Segunda estrella:** Este nivel cuenta con un formato estructura que sea legible por máquina, lo que permite al consumidor el análisis y explotación de los datos de manera automatizada con software propietario. Aún cuenta con bloqueo sobre las libertades completas del consumidor ya que encuentra una dependencia constante con el software propietario para la utilización de los datos.

- **Tercera estrella:** Se incorpora la utilización de formatos abiertos para la disponibilización de los datos. El consumidor ya no depende de software propietario para el análisis y explotación. Este nivel es el mínimo necesario para la apertura de los datos.
- **Cuarta estrella:** En este escalón se encuentran los formatos estandarizados por la W3C, como es el caso de RDF. Los datos se agrupan por temáticas y se los vincula para ser almacenados y gestionados por tecnologías pertinentes.
- **Quinta estrella:** En este escalón toma relevancia la interacción de los consumidores de los datos, que aportan un nuevo tratamiento sobre la información y contribuyen con datos faltantes o la corrección de los existentes.

Es importante notar que este no es el único esquema para calificar los datos abiertos. Por ejemplo, el Instituto de Datos Abiertos^[5] (ODI) también tiene un sistema de medallas con criterios similares pero dividido en cuatro niveles^[6]. Vale aclarar que alcanzar las cinco estrellas propuestas por Tim Berners-Lee no representa un nivel de calidad perfecta sobre los datos o que estos no puedan mejorar, ya que este esquema no evalúa la calidad propiamente dicha sino que describe una base de cómo se dispone la información a un ente consumidor. Para hacer hincapié sobre la calidad de los datos se deben seguir otros estándares como la ISO/IEC 25012^[7] que son guías relacionadas a las buenas prácticas de limpieza y verificación de los datos. La insistente búsqueda de la calidad de los datos^[8] por sobre la cantidad de los mismos lleva a determinar puntos de valor en sus características, entre ellas se encuentran:

- **Compleitud.**
- **Exhaustividad.**
- **Puntualidad.**
- **Oportunidad.**
- **Comparabilidad.**
- **Interoperabilidad.**

Además, normas ISO como la mencionada anteriormente aplican sobre los datos disponibles aún siendo de carácter abierto, esto quiere decir que pueden mantener las mismas características y nutrir la información disponibilizada.



Atributos de calidad propuestos por la norma ISO/IEC 25012.

Así mismo se encuentran diversas guías de calidad de los datos abiertos propuestas por diferentes organismos internacionales que se encargan de compartir este tipo de herramientas, pero son ellos mismos quienes incentivan sobre la generaciones de bases de calidad de los datos en cada una de las organizaciones vinculadas apoyándose en las ya existentes como en Europa (<https://www.europeandataportal.eu/elearning/en/module5/#/id/co-01>) y USA (<https://unstats.un.org/unsd/dnss/QualityNQAF/nqaf.aspx>), de allí surgen documentos que guía sobre las buenas prácticas y estructuras que deben tener los datos para ser disponibilizados en entidades gubernamentales como es el caso del Paquete de Apertura de Datos de la República Argentina^[9].

Instituciones estandarizadoras

El ecosistema de datos abiertos cuenta con una amplia cantidad de instituciones que enmarcan las estructuras, estándares y definiciones sobre los datos. Además, existen organizaciones especializadas en hacer posible la comunicación de las distintas regulaciones y políticas ejercidas sobre la apertura de datos. Entre ellas se encuentran:

- **Open Knowledge:** Es una organización internacional centrada en mostrar el valor de los datos abiertos a la sociedad, ayudando a asociaciones civiles a acceder y usar datos para reaccionar a problemas sociales. Trabaja sobre 3 núcleos que buscan mostrar el valor de los datos abiertos aplicados a las sociedades civiles, brindar herramientas y conocimientos para poder usar los datos y lograr que los sistemas de información de los gobiernos respondan a

la sociedad civil. Algunos de los proyectos más grandes del ámbito de los datos abiertos son liderados por la organización, como por ejemplo CKAN, una plataforma que permite el almacenamiento y distribución de datos abiertos; Frictionless data^[10], una herramienta que crea herramientas, estándares y mejores prácticas para la publicación de datos; Global Open Data Index^[11], que facilita el seguimiento del estado de los datos de gobierno abierto en todo el mundo; Open Data Handbook, que provee guías, casos de estudio y recursos para quien quiera publicar datos abiertos, teniendo en cuenta aspectos legales, sociales y técnicos, como también el por qué de la apertura, que significa que los datos estén abiertos y cómo es el procedimiento para abrirlos; o The Open Definition que define con precisión los principios de apertura en relación a los datos y al contenido, por lo que garantiza calidad y fomenta la compatibilidad entre diferentes conjuntos de datos abiertos. Esta es la principal norma internacional para los datos abiertos y licencias de datos abiertos.

- **Data Revolution Group:** Es un grupo de la Organización de las Naciones Unidas que impulsa acciones transformadoras como la producción y utilización de los datos; la disminución de las brechas para prevenir la discriminación, fomentar el desarrollo de la capacidad y la alfabetización de datos en datos pequeños y análisis de big data; la modernización de los sistemas de recopilación de datos; la apertura de datos para promover transparencia y el rendimiento de cuentas; y el desarrollo de nuevos objetivos e indicadores.
- **Open Government Partnership:** Como bien lo menciona en su página web, la Alianza de Gobierno Abierto, es una iniciativa multilateral que busca asegurar compromisos concretos de los gobiernos para promover la transparencia, capacitar a los ciudadanos, luchar contra la corrupción y aprovechar las nuevas tecnologías para fortalecer la gobernabilidad. En el espíritu de la colaboración de múltiples partes interesadas, OGP es supervisado por un Comité Directivo que incluye representantes de gobiernos y organizaciones de la sociedad civil. Fue creado en el año 2011, con 8 países que respaldaron la Declaración de Gobierno Abierto^[12] y anunciaron su plan de acción. Desde esa fecha se han sumado 67 gobiernos adicionales, lo que suma un total de 70 países y 15 gobiernos subnacionales con más de 2.500 compromisos asumidos en total.
- **Open Contracting Partnership:** La Alianza para las Contrataciones Abiertas es una organización que se encarga de facilitar la apertura de la contratación pública a través de la divulgación de los datos y el compromiso de apertura de información, de manera que las grandes sumas de dinero involucradas sean invertidas de forma honesta, justa y eficaz. La organización generó una estrategia para permitir que las contrataciones y las compras públicas pasen a ser de cerradas a abiertas, para esto crearon un estándar de datos para contrataciones abiertas^[13] que refleja el ciclo de contratación completo y permite que los usuarios y los socios de todo el mundo publiquen datos que se puedan compartir, reutilizar y sean legibles por máquinas. Su objetivo está en unir esos datos con su propia información y crear herramientas para analizarlos o compartirlos.
- **Open Data Institute:** Es una compañía independiente, sin fines de lucro y sin afiliación partidista fundada en el año 2012 con el fin de abogar por el uso innovador de datos abiertos para afectar el cambio positivo en todo el mundo.
- **Open Data Foundation^[14]:** Es una organización sin fines de lucro dedicada a la adopción de estándares de metadatos globales y al desarrollo de soluciones de código abierto que promueven el uso de datos estadísticos. Se enfoca en mejorar la accesibilidad de datos y metadatos, y la calidad general en apoyo de investigación, creación de políticas y

transparencia en los campos de economía, finanzas, salud, educación, trabajo, ciencias sociales, tecnología, agricultura, desarrollo y medio ambiente.

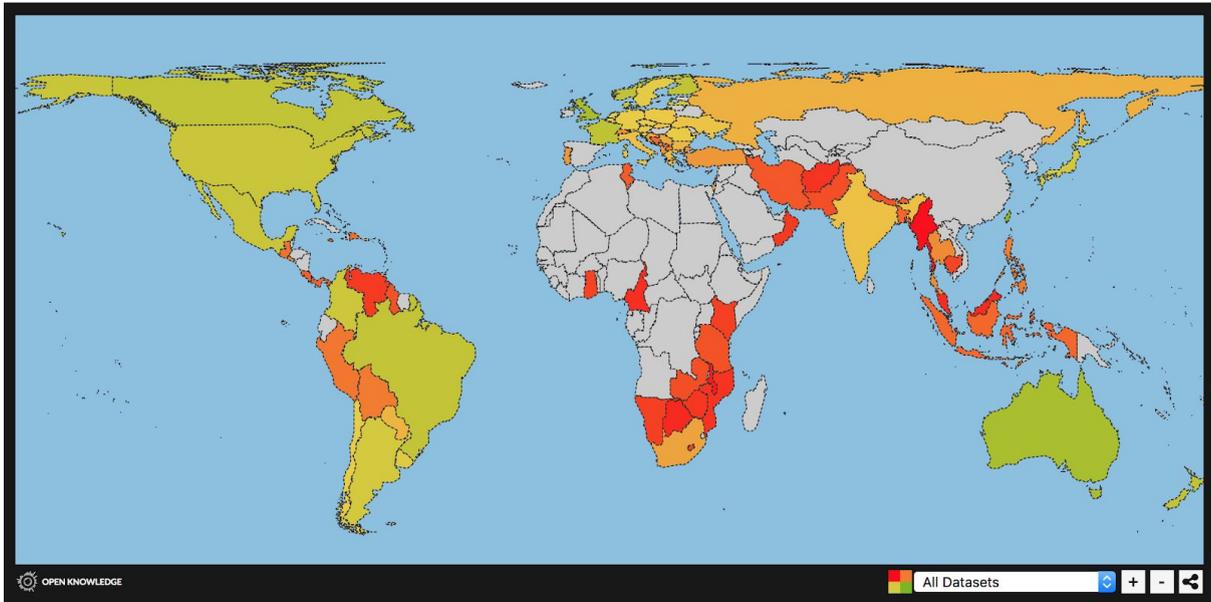
- **World Wide Web Consortium (W3C):** Es una comunidad internacional que se encarga de desarrollar los estándares web. Su misión es llevar la World Wide Web a su máximo potencial mediante el desarrollo de protocolos y directrices que aseguren el crecimiento a largo plazo de la Web. Entre los estándares definidos se encuentra el Data on the Web Best Practices^[15] donde se definen los estándares de uso de datos en la web.
- **Creative Commons^[16]:** Es una red de personas, juntas directivas y afiliados de todo el mundo, que ofrece licencias de copyright gratuitas y fáciles de usar con la finalidad de brindar al público permiso para compartir y utilizar un trabajo creativo de forma simple y estandarizada.
- **Iniciativa Latinoamericana de Datos Abiertos^[17]:** Conocido como ILDA, es un proyecto que busca entender y promover las políticas y usos de datos abiertos en América Latina para favorecer el desarrollo inclusivo de la región. La iniciativa forma parte de Open Data for Development^[18] como nodo regional para América Latina. Surge luego de la primera Conferencia Regional de Datos Abiertos en Montevideo, Uruguay, en 2013, con el apoyo del Centro de Investigaciones para Desarrollo de Canadá^[19] y tiene con objetivo principal investigar y colaborar con la naciente comunidad latinoamericana de datos abiertos y tecnologías cívicas. Además promueve investigaciones básicas y aplicadas sobre políticas y usos de datos abiertos, así como también asiste a gobiernos y sociedad civil en proyectos vinculados con este campo.
- **Sunlight Foundation^[20]:** Es una organización sin fines de lucro que se encarga de fomentar la participación ciudadana y establecer mecanismos con los cuales se logre un mayor grado de transparencia en el gobierno en cuestión. Para ello utiliza tecnologías cívicas, datos abiertos, el análisis de las políticas establecidas y, fundamentalmente, el periodismo como herramienta crítica para la exigencia de rendición de cuentas y el acercamiento de la información al ciudadano. Además, se encarga de generar informes de conflictos de interés, controlar el avance de las ciudades abiertas y sus iniciativas contra la corrupción y la democracia. Fue parte de la comisión que aprobó reformas históricas como la Ley de Libertad de Información^[21] (FOIA, por sus siglas en inglés), entre otras.

Métricas de los datos abiertos

Existen métricas sobre los datos abiertos que permiten llevar un control y analizar qué iniciativas y acciones se están llevando adelante por parte de los gobiernos para fomentar la transparencia y eficiencia en políticas públicas. Estas métricas visibilizan los vacíos de información y permiten identificar oportunidades para lograr que los datos sean más utilizables y tengan mayor impacto en los usuarios. Por otro lado, proporcionan a los gobiernos información sobre la calidad, consistencia y grado de apertura de sus datos. Entre los organismos que proporcionan métricas abiertas se encuentran:

- **Global Open Data Index:** Conocido como GODI, el Índice Global de Datos Abiertos es tomado anualmente como punto de referencia mundial para la publicación de datos de gobierno abierto. Su administración la realiza la Open Knowledge Network^[22] y mide la apertura de los datos gubernamentales de acuerdo con la definición abierta. Las variables

con más participación de los gobiernos son: presupuesto del gobierno, estadísticas nacionales y leyes nacionales.

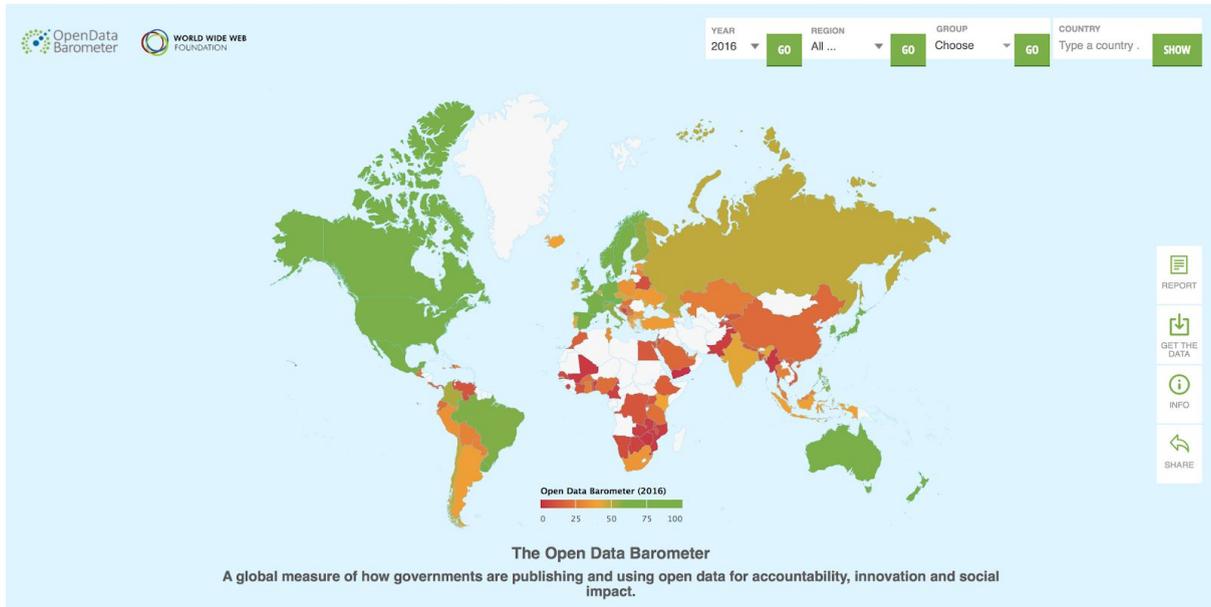


Representación gráfica de GODI.

Rank	Place	Government Budget	National Statistics	Procurement	National Laws	Administrative Boundaries	Draft Legislation	Air Quality	National Maps	Weather Forecast	Company Register	Election Results	Locations	Water Quality	Government Spending	Land Ownership	Score
1	Taiwan	Green	Green	Green	Red	Green	Green	Green	Green	Green	Green	Green	Green	Green	Red	Green	90%
2	Australia	Green	Green	Green	Red	Green	Red	Red	Green	Red	Green	Green	Green	Green	Red	Red	79%
2	Great Britain	Green	Green	Red	Red	Green	Red	Green	Green	Red	Green	Red	Red	Green	Red	Red	79%
4	France	Green	Green	Red	Red	Green	Red	Red	Red	Red	Green	Green	Red	Red	Red	Red	70%
5	Finland	Red	Green	Red	Red	Green	Red	Red	Green	Red	Red	Green	Green	Green	Red	Red	69%
5	Canada	Red	Green	Green	Green	Green	Red	Red	Green	Red	Green	Green	Red	Red	Red	Red	69%
5	Norway	Green	Green	Red	Red	Green	Red	Red	Green	Green	Green	Red	Green	Red	Red	Red	69%
8	New Zealand	Green	Red	Red	Red	Green	Red	Red	Red	Red	Red	Red	Red	Green	Red	Red	68%
8	Brazil	Green	Green	Red	Green	Red	Red	Red	Green	Red	Red	Green	Red	Green	Red	Red	68%
10	Northern Ireland	Red	Green	Red	Green	Green	Red	Green	Red	Green	Green	Red	Red	Green	Red	Red	67%

Representación tabular de GODI.

- Open Data Barometer (ODB)**^[23]: Es un proyecto colaborativo de la World Wide Web Foundation junto a Open Data for Development (OD4D)^[24] y Omidyar Network^[25]. Analiza tendencias globales y provee información comparativa de países y regiones utilizando una metodología que combina información contextual y evaluaciones técnicas. Posiciona a los gobiernos en un ranking según la disposición para llevar adelante iniciativas de datos abiertos, la implementación de programas de datos abiertos y el impacto que los datos abiertos tienen en los negocios, la política y la sociedad civil.

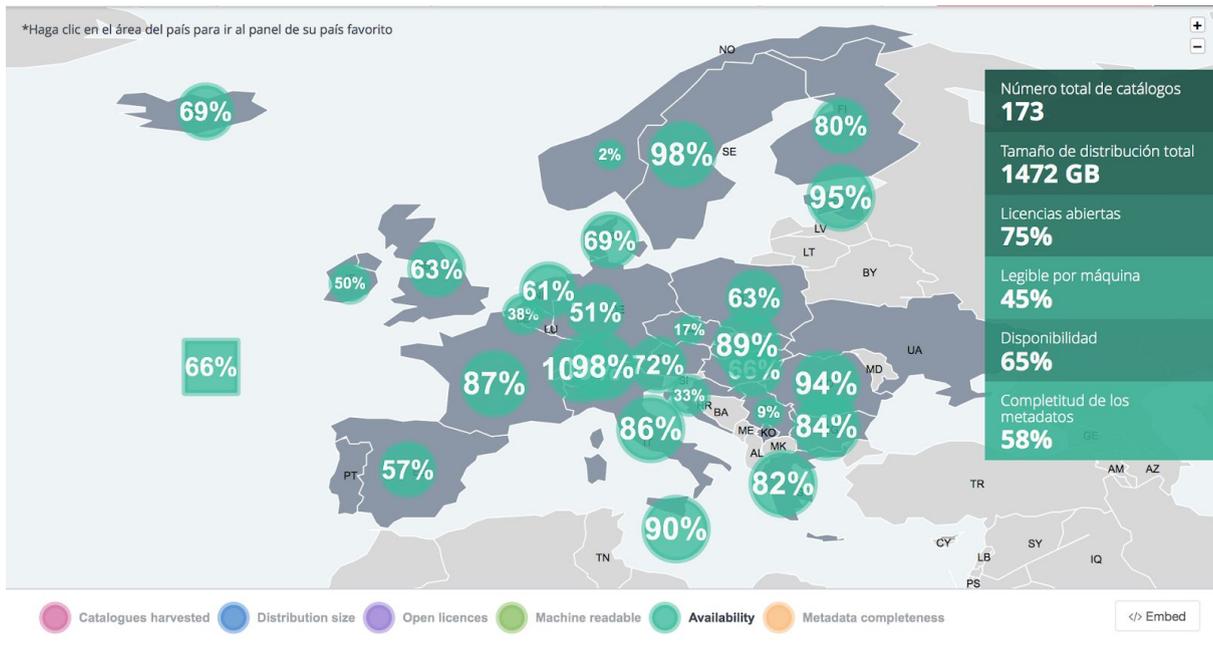


Representación gráfica de ODB.

Country	Rank ?	Score ? OUT OF 100	Change ?	Score Trend ? OVER PAST EDITIONS	Readiness ? OUT OF 100	Implementation ? OUT OF 100	Emerging Impact ? OUT OF 100
United Kingdom See details	1	100	0 -		99	100	94
Canada See details	2	90	2 ▲		96	87	82
France See details	3	85	-1 ▼		100	71	88
United States of America See details	4	82	-2 ▼		96	71	80
Korea See details	5	81	3 ▲		95	59	100
Australia See details	5	81	5 ▲		85	78	78
New Zealand See details	7	79	-1 ▼		92	58	99
Japan See details	8	75	5 ▲		84	60	89
Netherlands See details	8	75	-1 ▼		94	64	68
Norway See details	10	74	7 ▲		77	71	73

Representación tabular de ODB.

- Open Data Monitor^[26]**: Proyecto financiado dentro del Séptimo Programa Marco de Investigación y Desarrollo de la Unión Europea^[27], cuyo principal objetivo es investigar y desarrollar un esquema y metodologías para abordar la gran cantidad y diversidad de los metadatos disponibles de datos abiertos. Esto permite analizar y entender los metadatos, con el fin de comprender mejor los conjuntos de datos abiertos. Este proyecto se centra en entender mejor la situación de los datos abiertos en la región europea. Las variables tomadas en cuenta para hacer el análisis son la cantidad de conjuntos de datos con licencias abiertas, la cantidad de datos legibles por máquinas, la disponibilidad de los datos y la completitud de los mismos.



Representación gráfica de Open Data Monitor.

Pais	Número de conjunto de datos	Licencias abiertas	Legible por máquina	Disponibilidad	Completitud de los metadatos
Austria	2377	80%	28%	72%	69%
Belgium	3742	2%	59%	38%	36%
Bulgaria	103	11%	93%	84%	59%
Croatia	9	0%	100%	33%	69%
Czech Republic	44	0%	65%	17%	22%
Denmark	995	1%	53%	69%	57%
Estonia	41	25%	63%	95%	50%
Finland	1769	24%	49%	80%	79%
France	15653	31%	47%	87%	72%
Germany	28892	24%	34%	51%	55%
Greece	716	32%	78%	82%	62%
Hungary	162	28%	60%	66%	44%

Representación tabular de Open Data Monitor.

Es de esperar que estas organizaciones definan el rumbo que las instituciones deben seguir para lograr las mejores prácticas sobre la apertura de los datos, establezcan parámetros de control sobre las áreas involucradas y promuevan la competencia por lograr un mayor grado de excelencia en la iniciativa de datos abiertos.

Iniciativa de datos abiertos en el mundo

Cabe destacar que las iniciativas de datos abiertos a nivel mundial por parte de los países más desarrollados se encuentran vinculadas a áreas de investigación científica que fueron los puntos de partida para adentrarse con el paso de los años en áreas de mayor envergadura. Hoy en día, el

área gubernamental es la de mayor visibilidad en los datos abiertos. Si bien cada país comenzó con esta iniciativa en años diferentes, todos toman como base el Memorándum establecido por el ex presidente de los Estados Unidos, Barack Obama, como se mencionó en el capítulo anterior.

Para comenzar, Estados Unidos es uno de los dos países más preponderantes en lo que respecta a la iniciativa de datos abiertos. Las políticas establecidas por sus gobiernos persiguen el objetivo de fomentar el crecimiento económico y la creación de empleo. Dichas políticas están basadas en las definidas por el Sunlight Foundation y Open Knowledge, como se mencionó anteriormente, dos instituciones pioneras en el tema. Además de los objetivos enunciados, a partir de la iniciativa de gobierno abierto propuesta, se establecen las bases para lograr un gobierno eficiente, eficaz y transparente que fomente la participación pública y la colaboración por parte del ciudadano. A su vez, el gobierno de Estados Unidos creó un Proyecto de Datos Abiertos^[28] con el fin de compartir buenas prácticas, ejemplos y sistemas que permitieran a las demás entidades adentrarse en la apertura de los datos.

Otro país de gran relevancia respecto a la iniciativa de datos abiertos, es el Reino Unido. Es importante mencionar que uno de sus mayores logros fue la creación de la Licencia de Gobierno Abierto^[29] (OGL, por sus siglas en inglés Open Government License). Esta licencia permite que cualquier ente copie, publique, distribuya, transmita y adapte los datos licenciados y los explote con o sin fines de lucro. A cambio, dicho ente debe reconocer el origen de los datos y proporcionar un enlace a OGL. El gobierno del Reino Unido se vió ante la negativa de apertura de los datos por varios años, hasta que finalmente terminaron logrando la aplicación total de dicha política en el ámbito del transporte, que puso a disposición los horarios de los operarios ferroviarios y de autobuses utilizando formatos abiertos en cada una de sus plataformas.

Francia, es otra de las naciones que tomó gran importancia en la iniciativa de datos abiertos. Ya existían grandes instituciones no gubernamentales, como Wikimedia France^[30] y Open Street Map France^[31], entre otros, que disponibilizaban los datos de forma abierta; pero en lo que respecta a datos abiertos gubernamentales, el inicio se establece en 2010. Las leyes que rigen sobre la política de datos públicos abiertos surgen en la rendición de cuenta establecida en la primera constitución, la cual se fue perfeccionando en la Ley de Acceso a la Información Pública. Además, Francia brinda acceso a la información bajo dos licencias: Licence Ouverte^[32] y Open Database License^[33].

Un estado que ha mostrado gran crecimiento en la iniciativa de datos abiertos es Canadá, cuyas políticas orientaron a las instituciones públicas a perseguir un alto grado de perfección con la cual disponibilizar los datos. La diversa cantidad de áreas en las que Canadá se ve afectada por la apertura de datos la posiciona en un lugar destacado en la escala de países con mayor eficiencia en datos abiertos; entre las áreas que incursiona la iniciativa de datos abiertos en Canadá se encuentran el periodismo, transporte y áreas de georeferencia, entre otras.

Cabe destacar que las políticas llevadas a adelante por los diferentes países respecto de la apertura de datos de gobierno, marcan las reglas y/o regulaciones sobre distritos y/o ciudades que quedan relegadas en dicho aspecto por las demás. Establecen el punto de partida para gestionar las leyes que prevalecerán sobre la necesidad de disponibilizar la información con el fin de mejorar las instituciones gubernamentales, reducir los costos en la comunicación de los servicios y enriquecer los servicios públicos, entre otros. Este es el caso de ciudades que han avanzado de manera veloz en la iniciativa, como Madrid^[34], Toronto^[35], Chicago^[36], Nueva York^[37] o Londres^[38]. Todas ellas concentran una amplia variedad de proyectos, aplicaciones y plataformas que utilizan los datos abiertos recolectados de las diversas áreas de interacción explicadas en la siguiente sección, cuando se detallen los casos de éxito del uso de los datos.

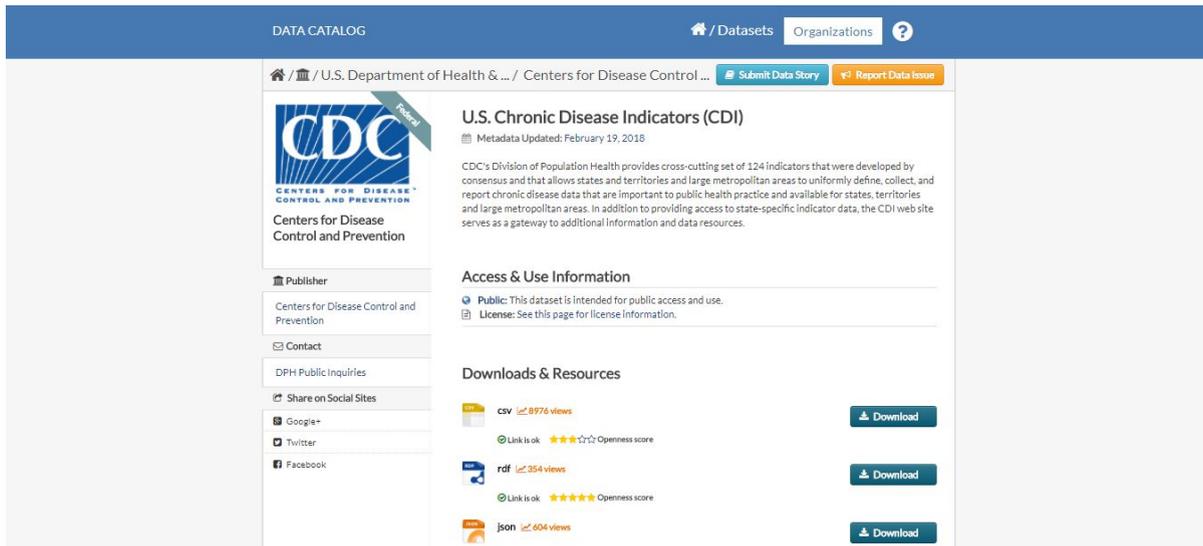
En este sentido, las políticas establecidas por los gobiernos que forman parte de la iniciativa de gobierno abierto llevan a cabo la generación de portales web con el fin de poner a disposición los datos gubernamentales de forma clara y concisa, legible por máquina y de manera intuitiva al consumidor. Si bien cada herramienta provee su estructura y formato en que se verán enmarcados los datos y los conjuntos de datos, todos se centran en categorías o áreas que nuclean la información y, sin hacer distinciones, los estándares resultan ser las áreas de salud, comercio, institucionales, educación, seguridad y medio ambiente; sujeto a las regulaciones o restricciones de cada país, ciudad o ente que utilice este tipo de plataformas. Muchos, a su vez, permiten las visualizaciones de los datos en forma más amigable para lograr una mayor aceptación por parte de la población, como suelen ser los gráficos estadísticos o mapas de georeferenciación.

Los portales de datos abiertos ganaron mucho terreno en los últimos años. Entre los más destacados se encuentran:

- **Estados Unidos** (<https://www.data.gov/>): este portal fue generado luego de las políticas establecidas en 2009. Utiliza la herramienta CKAN y Wordpress^[39] para la apertura de la información. Establece un canal especial para realizar solicitudes de conjuntos de datos por partes de entidades competentes, que son evaluados y resueltos por el equipo de datos en cuestión. Esta plataforma cuenta con 233.708 conjuntos de datos sobre los cuales se pueden hacer diferentes filtros, búsquedas, visualizaciones de gráficos o mapas, dependiendo del tipo de datos que se analice. Dentro de los formatos abiertos utilizados se encuentran json, rdf, csv y xml.

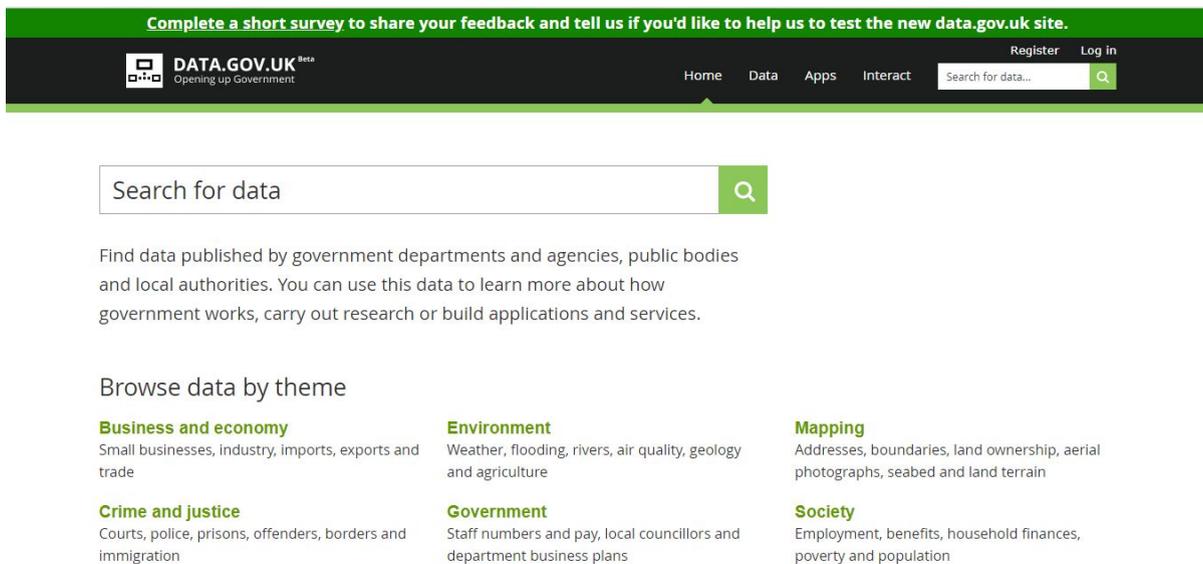


Página de inicio del portal de datos abiertos de Estados Unidos. Se pueden ver claramente las categorías que utiliza como filtro inicial sobre el conjunto total de datos.



Página de visualización de un conjunto de datos en particular donde se encuentran los recursos que conforman dicho conjunto en los formatos abiertos enunciados anteriormente.

- Reino Unido** (<https://data.gov.uk/>): este portal fue creado en 2010 basado en la herramienta CKAN y hoy en día cuenta con 44.419 conjuntos de datos. Disponibiliza información de una amplia variedad de áreas y posibilita la comunicación directa a través de blogs y comentarios dentro de los conjuntos de datos. Se encuentra bajo la licencia OGL 3.0. Utiliza los ya mencionados formatos abiertos para la disposición de los datos y permite previsualizaciones en algunos tipos de datos. Es importante destacar que cada conjunto de información cuenta con un grado de apertura dada las métricas y normas enunciadas anteriormente.



Página de inicio del portal de datos abiertos de Reino Unido. Se puede ver claramente las categorías que utiliza como filtro inicial sobre el conjunto total de datos.

Complete a short survey to share your feedback and tell us if you'd like to help us to test the new data.gov.uk site.

DATA.GOV.UK^{beta}
Opening up Government

Home Data Apps Interact Search for data... Register Log in

Datasets Map Search Data Requests Publishers Data API Organograms Site Analytics Reports Contracts

Home / Datasets / Road Safety Data

Road Safety Data

Published by Department for Transport. Licensed under **OGL** Open Government Licence.
Openness rating: ★★☆☆☆ Open Data Certificate: Bronze Level

These files provide detailed road safety data about the circumstances of personal injury road accidents in GB from 1979, the types (including Make and Model) of vehicles involved and the consequential casualties. The statistics relate only to personal injury accidents on public roads that are reported to the police, and subsequently recorded, using the STATS19 accident reporting form.

All the data variables are coded rather than containing textual strings. The lookup tables are available in the "Additional resources" section towards the bottom of the table.

Please note that the 2015 data were revised on the 29th September 2016.

Accident, Vehicle and Casualty data for 2005 - 2009 are available in the time series files under 2014. Data for 1979 - 2004 are available as a single download under 2004 below.

Also includes: Results of breath-test screening data from recently introduced digital breath testing devices, as provided by Police Authorities in England and Wales
Results of blood alcohol levels (milligrams / 100 millilitres of blood) provided by matching coroners' data (provided by Coroners in England and Wales and by Procurators...

Read More

DATA RESOURCES (48 IN A TIME SERIES)

2016 View Less

Página de visualización de un conjunto de datos en particular donde se encuentran los recursos que conforman dicho conjunto y se puede apreciar el grado de apertura mencionado anteriormente.

- Canadá** (<https://open.canada.ca/en>): este portal fue creado en 2011 y hoy en día cuenta con 79.094 conjuntos de datos. Se diferencia de las plataformas anteriores respecto de la estructura presentada en la navegación del sitio. Se encuentra orientado hacia el público que corresponde especialmente a datos. Disponibilizan los mapas y aplicaciones que surgen de la utilización de los datos disponibles. Cuenta además con el grado de apertura dentro del conjunto de datos específico.

Government of Canada / Gouvernement du Canada

Search Canada.ca

Jobs Immigration Travel Business Benefits Health Taxes More services

Home → Open Government → Search Open Government

Open Government Portal

Search Records

Search... Suggest a Dataset

79,094 records found Order by Relevance

Drug Product Database - All Files

The Drug Product Database (DPD) system captures information on Canadian human, veterinary and disinfectant products approved for use by Health Canada. This extract contains both marketed (active) and discontinued (inactivated) products in separate files. This data extract contains...

Organization: Health Canada

Search Filters Clear All

Portal Type

- Open Data (78819)
- Open Information (275)
- Clear All

Collection Type

- Geospatial (69487)
- Non-Spatial (8573)
- Open Maps (759)
- Publications (275)

Página de inicio del portal de datos abiertos de Canadá. Se puede ver claramente las categorías que utiliza como filtro inicial sobre el conjunto total de datos.

Government of Canada / Gouvernement du Canada

Search Canada.ca

Jobs | Immigration | Travel | Business | Benefits | Health | Taxes | More services

Home → Open Government → Drug Product Database - ...

Drug Product Database - All Files

The Drug Product Database (DPD) system captures information on Canadian human, veterinary and disinfectant products approved for use by Health Canada. This extract contains both marketed (active) and discontinued (inactivated) products in separate files. This data extract contains information about Drug Product - All Files.

Publisher - Current Organization Name: Health Canada

Licence: Open Government Licence - Canada

Resources

Resource Name	Resource Type	Format	Language	Links
Active Ingredients	API	JSON	English	Access

Have your say

Rate this dataset
16 Comment(s)

Additional Information

Contact Email: open-ouvert@tbs-sct.gc.ca

Keywords: [dpd](#) [drug product](#) [pharmaceutical drugs](#) [biological drugs](#) [brand name](#)

Página de visualización de un conjunto de datos en particular donde se encuentran los recursos que conforman dicho conjunto y se puede apreciar el grado de apertura mencionado anteriormente.

En conclusión, las leyes y políticas establecidas por los gobiernos, los procesos de análisis y selección de los datos detallados en la sección anterior, y los portales desarrollados como soporte tecnológico de las iniciativas de datos abiertos, conforman los pilares de gobierno abierto y generan un valor agregado a las distintas áreas que lo constituyen.

Iniciativa de datos abiertos en Latinoamérica y Argentina

Ante la inminente aparición de la iniciativa de datos abiertos a nivel mundial en los últimos diez años, es importante mencionar el gran avance que han tenido los países latinoamericanos en cuanto a las políticas desarrolladas para permitir un mayor grado de apertura de la información de cada una de las organizaciones que interactúan directa o indirectamente con sus gobiernos. No es menor la brecha que existe entre los países desarrollados y los que se encuentran en un escalón inferior, no solo en materia de datos abiertos, sino también en el aspecto tecnológico. Esta impronta que lleva adelante cada uno de los gobiernos de América Latina es el pilar de un paradigma en la región que tiene como objetivo hacer frente a la corrupción, al estado en que se encuentran los servicios públicos y los costos exuberantes que aparecen en contrataciones y servicios prestados.

Es importante destacar que la región cuenta con un bajo grado de apertura de la información, que está teniendo sus comienzos en cada una de las áreas y que recién se está incursionando en los beneficios que conlleva ser un gobierno abierto. Es indudable que las políticas que cada gobierno establezca se deben llevar adelante con sumo cuidado y consenso entre los organismos ya que mucho de ellos todavía se rehúsan a disponibilizar información. Por este motivo, no es menor resaltar la decisión de los gobiernos latinoamericanos de establecer nuevas leyes y desafíos que persiguen dicha iniciativa.

Uno de los organismos con mayor contribución en la región es ILDA, Iniciativa Latinoamericana por los Datos Abiertos, cuya principal labor en los últimos cinco años es enmarcar los ejes claves de una buena estrategia de apertura de datos para que los gobiernos se apropien de las iniciativas correspondientes y logren la eficacia y eficiencia en sus organizaciones. ILDA actúa como facilitador de espacios de consenso y reunión, aporta un panorama amplio de capacitaciones

sobre las ventajas de incentivar la apertura de la información, genera informes sobre los avances que se presentan en dicha área basándose en los desafíos propuestos por cada uno de los gobiernos y establece la conexión entre los representantes de gobierno abierto de cada uno de los países involucrados.

Uno de los países que se encuentra mejor posicionado en lo que respecta a datos abiertos, es México^[40]. Posee un avanzado esquema de leyes que rigen sobre la disponibilización de los datos abiertos y cómo deberían aportar cada una de las organizaciones a este tema. Si bien cuentan con un amplio conjunto de datos de varias áreas de gobierno, algunas carecen de detalle. De todas formas cabe destacar el alto impacto generado por el uso de datos abiertos, ya sea por plataformas externas como por aplicaciones en tiempo real. Uno de los avances más notorios en este país es el de contrataciones abiertas, las cuales hacen transparente todo tipo de convenios y contrataciones públicas.



Plataforma de Contrataciones Abiertas del Gobierno Mexicano^[41].

Brasil, que se encuentra en uno de los escalones más altos en la consideración de Open Data Barometer a nivel regional, fue uno de los primeros países latinoamericanos en incursionar en la transparencia de los sectores financieros. En 2004 se creó el Portal de Transparencia^[42] con el fin de hacer visible el estado fiscal del Gobierno Federal brasileño a través de los datos del presupuesto del gobierno. Con el correr de los años, el portal ha crecido de manera exponencial en cuanto a la calidad y cantidad de datos disponibilizados. Hoy en día es la herramienta principal que tiene el gobierno para hacer frente a la corrupción. Esta plataforma fue pionera y sirvió de modelo a varios de los países de Latinoamérica para encontrar el camino adecuado hacia las iniciativas de gobierno abierto.

Otro país que ha avanzado notoriamente en temas de gobierno abierto es Colombia que cuenta con una sólida estructura de leyes que impactan sobre la iniciativa de datos abiertos, una considerable planificación sobre la temática abordada de los datos a disponibilizar y una creciente plataforma interactiva con la cual ponen al corriente al ciudadano. Uno de los aspectos a destacar es la búsqueda de concientizar a los potenciales consumidores de los datos sobre las nuevas políticas públicas y el impacto que generan con su utilización; esta estrategia tiene la finalidad de captar mayores comunidades que incursionen en nuevos conjuntos de datos, desarrollar nuevas plataformas o aplicaciones que beneficien los servicios públicos y generar eventos de difusión siguiendo estos nuevos lineamientos. Su plataforma de datos abiertos (<https://www.datos.gov.co/>)

cuenta con una estructura intuitiva y fácil de utilizar para quien no cuenta con avanzados conocimientos sobre datos abiertos, lo cual la hace más interesante para el ciudadano. Una característica para destacar es la disponibilización de los datos mediante una visualización, ya sea un gráfico o un mapa, y que la misma puede ser generada por el usuario de la plataforma de modo interactivo.

Chile, por su parte, ha avanzado paulatinamente sobre las leyes que fomentan los datos abiertos y en consecuencia se observa un gran compromiso por parte de las áreas que abarcan este tema para lograr los objetivos de gobierno abierto. Hoy en día, cuenta con una plataforma (<http://datos.gob.cl/>) dedicada a la comunicación directa con el ciudadano para poder obtener sugerencias, realizar informes sobre los lineamientos establecidos por su gobierno y ejercer conciencia sobre las políticas de gobierno abierto. Una de las plataformas a destacar es Datos Abiertos Chile Compra^[43] donde se encuentran los importes y modalidades de contratación (convenios, licitaciones, etc) de las compras realizadas por el gobierno nacional. Un agregado es que se pueden visualizar los consumos históricos en detalle.

Argentina, por su parte, se encuentra un escalón por debajo de estos países en lo que respecta a las políticas establecidas y leyes aplicadas sobre datos abiertos. En 2012, el país se unió a la Alianza para el Gobierno Abierto (OGP) con el fin de incursionar en la iniciativa de gobierno abierto y alcanzar los desafíos propuestos por el gobierno nacional sobre mejorar la calidad de la comunicación entre las distintas áreas de gobierno, modernizar los servicios prestados y hacer frente a los niveles bajos de transparencia en sus instituciones. En los últimos años, se ha fomentado esta iniciativa estableciendo la Ley de Acceso a la Información Pública^[44], sancionada en septiembre de 2016, que obliga a cualquier ente que perciba aportes públicos a responder las solicitudes de información que realice cualquier ciudadano. Cabe destacar que este esquema a nivel nacional no cuenta con un alto grado de madurez, pero existe un notorio énfasis en lograr una mejora sustancial sobre las organizaciones nacionales y poder establecer un punto de partida para las organizaciones y entes provinciales y municipales. Su plataforma de datos abiertos (<http://datos.gob.ar/>) presenta funcionalidades acotadas; si bien cuenta con una gran cantidad de conjunto de datos, no proporciona visualizaciones ni permite el contacto directo con el ciudadano.

Las corrientes de datos abiertos en Argentina se encuentran más avanzadas a nivel municipal y en gran cantidad de organismos privados que dieron impulso por cuenta propia sobre dicha iniciativa tomando de ejemplo otros países más avanzados en el tema. La Ciudad Autónoma de Buenos Aires cuenta con un alto grado de avance sobre esta iniciativa. Fue una de las primeras en gestionar áreas especializadas en gobierno abierto y en disponibilizar una plataforma ágil e interactiva para el ciudadano. Hoy en día, la plataforma (<https://data.buenosaires.gob.ar/>) cuenta con una gran cantidad de datos de todas las áreas de gobierno y establece una rápida comunicación con el ciudadano dándole la posibilidad de sugerir información y disponibilizando visualizaciones de tipo gráfico o geográfico según el dato específico. No solo se debe mencionar a la Ciudad Autónoma de Buenos Aires por su portal de datos abiertos sino también por su constante foco en la organización de jornadas de participación ciudadana como hackatones, datatones o ideatones para potenciar la co-creación y fomentar la creación de diversas aplicaciones, investigaciones y prototipos, utilizando datos abiertos. Otro punto interesante a mencionar es el uso de los datos abiertos dentro del gobierno para el desarrollo de aplicaciones que facilitan la interacción de los ciudadanos con los servicios públicos como “BA Cómo Llego”^[45], “BA EcoBici”^[46] y “BA Subte”^[47].

Bahía Blanca es otro de los municipios argentinos que se encuentra avanzado en la materia de datos abiertos. Dispone de una gran cantidad de datos que corresponden a una amplia variedad de áreas de gobierno y que cumplen los estándares necesarios para el consumo de la información a través de lenguaje de máquina. Su plataforma (<http://datos.bahiablanca.gob.ar>) permite la

visualización de los datos mediante gráficos o mapas, según el tipo de dato, y una intuitiva comunicación con el ciudadano. Es interesante destacar que dicha plataforma de datos abiertos tiene la capacidad de generar conjuntos de datos en tiempo real, preferentemente de los distintos sensores que se encuentran ubicados en puntos estratégicos del municipio para diversos fines y que pueden visualizarse a modo de gráfico. Además, dicho gobierno entrega información georeferenciada sobre los distintos casos de 911 emergencias para lograr disponer al ciudadano de un mapa de calor de las áreas con mayor incidencia. Otra de las plataformas interesantes de este municipio es Hospital Abierto^[48], que tiene como objetivo transparentar los recursos del hospital principal de dicho municipio y promover la iniciativa a nivel nacional.

La provincia de Santa Fe de la República Argentina dispone de un portal de datos abiertos (<https://www.santafe.gob.ar/datosabiertos/>) en el que se distinguen grandes volúmenes de datos correspondientes a áreas como salud, finanzas y educación entre las más destacadas. Dicha plataforma se encuentra desarrollada en Wordpress y DKAN, y cuenta con la limitante de que no se visualizan los datos en forma de gráficos o mapas como se ha visto en otros portales. Uno de sus puntos fuertes es la fácil interacción con el ciudadano para que sugiera y comunique nueva información o realice consultas sobre los datos. Además cuenta con una sección exclusiva de Presupuesto Abierto en la que figuran, a modo histórico, los gastos e ingresos de cada sector gubernamental en conjunto con los objetivos y las finalidades de los mismos. Además, el gobierno de Santa Fe a través del laboratorio de innovación ciudadana conocido como Santalab^[49], nuclea la colaboración entre el Estado, la ciudadanía, las organizaciones y las empresa, con el fin de co-gestionar proyectos de innovación a partir de los cuales se haga uso y generación de nuevos contenidos vinculando datos abiertos. La ciudad de Rosario también cuenta con un portal de datos abiertos (<http://datos.rosario.gob.ar/>) que pone a disposición un gran conjunto de datos correspondientes a áreas estratégicas como movilidad, educación y cultura. Además, muestra un balance financiero y planes económicos que pueden ser visualizados en un gráfico interactivo. La plataforma permite acceder a una sección exclusiva donde se encuentra el contenido de diversas aplicaciones creadas a partir de los datos abiertos disponibilizados y desde la cual se puede acceder a las respectivas plataformas de descarga, ya sea para dispositivos Android o IOS. Entre estas aplicaciones se destacan: ¿Cómo Llego?^[50], que permite establecer un recorrido georeferencial dado un punto de origen y un punto de destino; y Agenda Cultural^[51], que disponibiliza un cronograma de los eventos culturales que se realizan en la ciudad, detallando su ubicación y una breve descripción de los mismos.

Es importante destacar que si bien los países de latinoamérica se encuentran por debajo de los países europeos y de norteamérica, se ubican en un escalón medio a nivel mundial en relación a las políticas e iniciativas de apertura de información cada vez más consolidadas. Tanto los consensos como los desafíos planteados por cada uno de los gobiernos son los pilares para un próspero progreso sobre la transparencia de las instituciones, las iniciativas de anticorrupción y la mejora en los servicios públicos.

Casos de éxito

Ante la enumeración de los objetivos y la gran variedad de información con que se encuentran las diferentes organizaciones gubernamentales, es necesario incluir el concepto de big data que fue mencionado en el capítulo anterior. El gran volumen de datos públicos tiene como consecuencia positiva la toma de decisiones precisas, consensuadas y evaluadas sobre información concreta y detallada. Las estadísticas y patrones que provee la evaluación y análisis de estos datos logra, en gran medida, establecer nuevas políticas y propuestas a mediano y largo plazo.

Hablar de datos abiertos implica hacer foco en el requerimiento de que éstos sean fácilmente legibles tanto por humanos como por máquinas. Si se cumple éste requisito y además los datos publicados son ricos en información y útiles, cualquier persona o institución que lo desee puede desarrollar proyectos sobre ellos. A continuación se mencionan proyectos de gran impacto que fomentaron réplicas en otras partes del mundo con la utilización de datos abiertos provistos por los gobiernos.

- **Transport for London (TfL)^[52]:** Es el organismo del gobierno local responsable de implementar y administrar la estrategia y los servicios de transporte en la capital del Reino Unido. Supervisa casi todos los aspectos del transporte en la ciudad, con 24 millones de viajes realizados a través de la red de transporte de Londres todos los días. TfL pone a disposición 62 conjuntos de datos separados. Se trata de una combinación de feeds en tiempo real (como tableros de salida de metros, interrupción del tráfico en vivo, llegadas de autobuses en vivo y la API de Journey Planner de TfL^[53]), conjuntos de datos fijos (como horarios, ubicaciones de estaciones e instalaciones de estaciones) y datos administrativos orientados a la transparencia (que detalla el rendimiento operativo, la remuneración de los directores, etc.). Es posible concluir que si TfL hubiera tomado la decisión de conservar sus datos y desarrollar todas sus aplicaciones internamente, tendría que haber superado los costos de desarrollo de más de £ 43 millones para entregar las 362 aplicaciones que actualmente funcionan con TfL abierto.
- **Dónde - Fundación Huésped^[54]:** Es una organización argentina con alcance regional que, desde 1989, trabaja en áreas de salud pública con el objetivo de garantizar el derecho a la salud y el control de enfermedades. En 2013 lanzó una plataforma de código abierto llamada "Dónde"^[55] cuyos objetivos son favorecer el acceso a servicios para el cuidado de la salud sexual y reproductiva, y promover la participación ciudadana para mejorar la calidad de los mismos. Entre las fuentes de datos abiertos que consume este proyecto se encuentran las brindadas por el Gobierno de la Ciudad de Buenos Aires, el Ministerio de Desarrollo Social de La Nación y el Ministerio de Salud de la Provincia de Buenos Aires entre otros.
- **Dengue - Paraguay:** Esta investigación es un ejemplo de cómo los datos abiertos pueden ser utilizados para combatir enfermedades. El Departamento Nacional de Vigilancia Sanitaria de Paraguay abrió los datos relacionados con la morbilidad del dengue, enfermedad endémica en la zona desde 2009. Aprovechando esta información, el investigador Juan Pane y su equipo de la Facultad Politécnica de la Universidad de Asunción, crearon un sistema de alerta temprana que puede detectar brotes de dengue con una semana de anticipación. Este modelo basado en datos, funciona siempre y cuando haya datos disponibles sobre morbilidad, clima y agua de la región.
- **Mejora Tu Escuela:** Es una plataforma^[56] en línea independiente y sin fines de lucro que pone a los padres como principales colaboradores para la mejora de la educación escolar. Además les brinda información sobre el desempeño de las escuelas, les permite comparar instituciones, y calificar y dar opinión sobre diferentes temas. También proporciona a los administradores de cada establecimiento, legisladores y ONGs información para identificar áreas que requieren mejoras y focos de corrupción. Todos estos puntos contribuyen en el proceso de aumento de la calidad general de la educación en México. Fue fundada en 2013 por el Instituto Mexicano para la Competitividad (IMCO) con el apoyo de Omidyar Network.
- **A tu servicio^[57]:** Es una plataforma web uruguaya creada por la sociedad civil de datos abiertos, Datos Abiertos, Transparencia y Acceso a la Información (DATA) en asociación con

el Ministerio de Salud de Uruguay. En base a los datos abiertos gubernamentales de salud y la ubicación del usuario, permite comparar proveedores locales de atención médica, basados en una amplia gama de parámetros e indicadores, entre los que se encuentran el tipo de instalación, la especialidad médica, el tiempo de espera, la cantidad de usuarios, los derechos de los usuarios, etc. Las comparativas que presenta la plataforma son muy intuitivas y simples de usar, ya que representan gráficamente las diferencias entre las variables.

- **Estacionamiento en tiempo real**^{[58][59]}: El avance de las tecnologías permitió que las ciudades comiencen a desarrollar proyectos de IoT como la sensorización en tiempo real de diferentes variables. Estos dispositivos generan un gran caudal de información catalogado como big data. Si estos datos son disponibilizados en formato abierto pueden surgir proyectos realizados por terceros, como por ejemplo el realizado en la ciudad francesa Issy-Les-Moulineaux, donde se creó un mapeo en tiempo real sobre la disponibilidad de estacionamientos en la vía pública.
- **Where does my money goes?**^[60]: Es un proyecto del año 2007, cuya idea original fue de la Open Knowledge Foundation que tiene su objetivo en lineamiento con la transparencia en la gestión pública. El objetivo central es ayudar a los ciudadanos a entender dónde se gasta el dinero público. Particularmente, deja ver a los ciudadanos de qué forma el gobierno está invirtiendo y utilizando lo recaudado en los impuestos a través de análisis y visualizaciones.
- **Properati**^{[61][62]}: Es un sitio de búsqueda de inmuebles en Latinoamérica que utiliza datos abiertos disponibilizados por los gobiernos locales y organizaciones vinculadas al rubro para gestionar mapas con las propiedades en alquiler y en venta, y permitirle a los usuarios realizar diferentes filtros a partir de los parámetros obtenidos de esos datos de terceros. A su vez, pone a disposición su propio conjunto de datos con información agregada como precios aproximados en moneda extranjera, descripción e imágenes adjuntas del inmueble. El valor agregado de la plataforma es la generación de reportes a partir de los datos administrados y que son disponibilizados de manera abierta, que involucran las contrataciones y precios promedio de los inmuebles de forma mensual. El equipo de Properati, además, genera diversos proyectos a partir de sus datos que valen la pena destacar, entre ellos se encuentran: Barrios argentinos en que las mujeres sufren más violencia de género^[63]; El valor del precio por manzana en Chile^[64]; El ahorro necesario de una pareja en Chile^[65]; Zona sísmica cerca de tu propiedad^[66]. Además, cuenta con una herramienta conocida como Heatmap^[67] que permite al usuario realizar mapas de calor teniendo en cuenta ciertas característica a la hora de elegir una zona para su nueva propiedad, como puede ser la probabilidad de inundación, que cuente con cierta especie de árbol en la zona o la cercanía al subte o metrobús, entre las más destacadas.
- **Mapa de potenciales villas y asentamientos**^[68]: Es un proyecto independiente en Argentina que surge producto de una Tesis de Posgrado de la Universidad de Buenos Aires a cargo de Federico Baylé. El mismo consiste en un mapa interactivo en el cual se exponen conocimientos de Machine Learning para la generación de probabilidades de que una zona territorial pueda ser villa determinada por ciertos datos y características. Para la misma, utilizó los datos disponibilizados del censo realizado por la ONG Techo^[69] en 2013. La característica más relevante para la determinación de un establecimiento como villa es la precarización de las condiciones de la vivienda; además se agrega el contexto territorial como puede ser disposición de cloacas, energía eléctrica o red de agua corriente. Es importante destacar que dado este esquema inicial de datos se pudo extender los conocimientos a través de Machine Learning^[70] para disponer un mapa nacional completo en

cuanto a las probabilidades de potenciales villas y asentamientos.

Desafíos en la región

Es de esperar que los desafíos que los gobiernos latinoamericanos establezcan sean numerosos debido a la mencionada brecha que existe en aspectos generales sobre las iniciativas de datos abiertos logradas por los países europeos y norteamericanos. Entre los desafíos^[71] que ya fueron consensuados por cada uno de los gobiernos latinoamericanos en congresos realizados por las distintas organizaciones internacionales dedicadas a la iniciativa de gobierno abierto, se encuentran:

- **Presentar los datos con un lenguaje simple y claro:** si bien varias de las áreas que interactúan con el gobierno cuentan con una correcta disponibilización de los datos, queda mucho tramo por recorrer en relación a dicho aspecto ya que no todas las áreas cuentan con equipos específicos de datos abiertos que permitan la normalización y análisis de la información a derivar.
- **Fortalecer la comunidad para la cooperación y el intercambio de conocimiento y experiencias:** hoy en día la comunidad de datos abiertos es acotada debido a que la iniciativa tomó impacto en la región recién hace unos pocos años. Por ello, agilizar los procesos de capacitación sobre áreas específicas de las organizaciones vinculadas a gobierno dará valor agregado sobre el ecosistema de datos abiertos.
- **Impulsar el emprendimiento e innovación con los datos:** es de suma importancia la generación de nuevas herramientas para mejorar los servicios y la comunicación entre los organismos y el ciudadano.
- **Fomentar un Estado anticorrupción:** uno de los aspectos más relevantes en la agenda de los gobiernos es lograr un grado de transparencia dentro de sus instituciones que permita eliminar la corrupción instalada en sus mecanismos.
- **Capacitar sobre el uso y el impacto de los datos abiertos:** se persigue la búsqueda de concientización sobre los datos abiertos (definición, uso, impacto y beneficios) hacia el ciudadano, para que este pueda ser partícipe de una mejora continua de un gobierno abierto.

Además, se debe hacer hincapié en el creciente interés por parte de los gobiernos latinoamericanos por incursionar en lo que se conoce como “contratos abiertos”. Si bien ya existen países como México y Paraguay que tienen avanzado este tema, varios se encuentran lejos de lograr esos niveles de apertura de la información. Sin embargo, este desafío se encuentra más presente en las agendas políticas de dichos gobiernos, sobre todo en el gobierno nacional.

- **Desafíos México:** Al ser el país más avanzado en Latinoamérica en el ámbito de datos abiertos, los desafíos planteados son puntuales y ricos en su avance. Uno de los puntos en los que se encuentra trabajando en conjunto con el Banco Interamericano de Desarrollo (BID) y la Carta Internacional de Datos Abiertos, consta de un ensayo donde se probará la implementación de la Guía de Datos Abiertos Anticorrupción. Por otro lado, con el fin de promover la participación del sector privado y de la sociedad civil en la materia de los datos abiertos, se lanzará una colaboración entre la plataforma del gobierno (datos.gob.mx) y la plataforma de la sociedad civil (<http://datamx.io/>) manejado por Codeando México. Por último,

se dará inicio a la segunda etapa de "Labora"^[72], la iniciativa llevada adelante por Estrategia Digital Nacional de la Presidencia de la República de México, Open Data Institute, Demos y la Embajada del Reino Unido en México; tiene foco en el desarrollo económico en base a datos abiertos y está compuesta por una plataforma para emprendedores cívicos y sociales que ofrece herramientas y conecta con empresas, mentores e inversionistas y acelera el impacto a través de la innovación con datos; en esta segunda etapa se pondrá mayor énfasis en apoyar el trabajo de incubadoras y aceleradoras sectoriales, sin dejar de crear capacidades de uso en PyMEs y en servidores públicos.

- **Desafíos Colombia^[73]:** Los principales desafíos que plantea este país son fortalecer y profundizar la democracia, fomentar la confianza de los ciudadanos en el Estado, generar desarrollo económico incluyente, construir paz territorial sostenible y convivencia pacífica, garantizar los derechos de los ciudadanos, y luchar contra la corrupción. En concreto algunos de los compromisos tomados por el sector ejecutivo fueron la prevención de violencia sexual en mujeres y niñas, la participación ciudadana y control social, la democracia ambiental, la gestión de las entidades públicas, y la promoción del empleo público. El sector del Senado de la República, particularmente la rama Legislativa, se compromete a garantizar acceso a la información y participación para influir en la toma de decisiones legislativas. Por otro lado, desde la Procuraduría General, se propone fortalecer la participación regional y juvenil y el diseño del Sistema de Alertas Tempranas (SAT) en corrupción y mala administración.
- **Desafíos Chile^[74]:** Las principales temáticas acordadas entre el gobierno y la sociedad civil que establecen los próximos compromisos de este país en el marco de la OGP, son cuatro:
 - **Neo-extractivismo, recursos naturales y medioambiente:** dentro de ésta categoría se encuentran compromisos como promover estándares de calidad e integración de la ciudadanía en la gestión de la energía y contribuir a la equidad y democracia ambiental.
 - **Políticas de protección social y políticas educacionales:** Los desafíos se centran en el objetivo de construir y poner a disposición de las escuelas herramientas para que generen sus planes de formación ciudadana. Por ejemplo, buscan contribuir con contenidos y recursos educativos abiertos en los establecimientos educativos y desarrollar e implementar un modelo de trabajo para potenciar el uso del lenguaje claro en la administración pública a través de recursos metodológicos.
 - **Modernización del Estado:** incluye proyectos como la definición de la política de datos abiertos en la administración del Estado según los principios de la carta internacional de datos abiertos; fomentar el uso del portal de datos abiertos (<http://datos.gob.cl>), como también el aumento, reutilización y calidad de los datos que lo componen; promover y desarrollar el uso de datos abiertos de las compras públicas como medio para fortalecer la transparencia del Sistema de Compras Públicas y mejorar la información territorial publicada en el Geoportal de Chile.
 - **Integridad y transparencia en la función pública:** en este eje, el principal compromiso corresponde a mejorar los estándares de transparencia a nivel de los gobiernos locales a través de la implementación de nuevas herramientas de gestión.
- **Desafíos Brasil^[75]:** Cinco temas fueron tomados como prioridad en la asunción de compromisos, ellos son:
 - **Mecanismos de participación de los ciudadanos:** consolidación, mejora y expansión de los mecanismos de participación ciudadana
 - **Transparencia de los fondos públicos:** mejora de los mecanismos de transparencia sobre los fondos de la Administración Pública Federal.

- **Fomentar un gobierno abierto en estados y municipios:** Difusión y apoyo a prácticas de gobierno abierto (transparencia, participación, rendición de cuentas e innovación) en gobiernos subnacionales.
 - **Innovación y gobierno abierto en educación:** Soporte tecnológico y recursos educativos para mejorar la enseñanza y el aprendizaje.
 - **Datos abiertos y transparencia activa en el medio ambiente:** Fomentar la divulgación de datos abiertos y la actualización de mecanismos activos de transparencia relacionados con el área ambiental.
- **Desafíos Argentina^[76]:** Los compromisos se encuentran divididos en cuatro categorías agrupados por las siguientes temáticas:
 - **Transparencia:** Algunos de los compromisos tomados en materia de transparencia son sistematizar y publicar información sobre los procesos de selección de magistrados, como también información sobre procesos disciplinarios sobre los mismos; acercar el presupuesto al ciudadano; difundir el proyecto de ley de presupuesto y monitoreo de políticas públicas prioritarias (<https://www.presupuestoabierto.gob.ar>), portal de datos productivos y plataforma sobre cambio climático; implementar el Estándar de Datos para las Contrataciones Abiertas (EDCA); disponibilizar en formato abierto información sobre equidad de género en ciencia y tecnología, y datos del Congreso Nacional,
 - **Rendición de cuentas:** En ésta temática podemos encontrar compromisos como la promoción de una reforma del régimen de financiamiento de los partidos políticos elaborada participativamente; el monitoreo del Plan Nacional de Acción para la Prevención, Asistencia y Erradicación de la Violencia contra las Mujeres; y recomendaciones de auditorías en cárceles,
 - **Participación ciudadana:** Los objetivos están centrados en la realización de capacitaciones en prácticas de Gobierno Abierto; la elaboración de manera participativa de un Plan de Igualdad de Oportunidades (PIO) en términos de género; la promoción de mecanismos de vinculación con la sociedad civil en el ciclo de auditoría; y la apertura de espacios de debate y capacitación sobre el proceso electoral en Argentina,
 - **Tecnología e innovación:** Por último y con menos compromisos se encuentra esta temática que abarca generar una plataforma interactiva centralizada en donde cualquier persona pueda relevar, informar, detectar y retroalimentar políticas públicas destinadas a jóvenes de entre 15 y 24 años en todo el territorio nacional; y contribuir a la transparencia y a la innovación electoral a través del uso de la tecnología en las actividades de información y capacitación electoral.
 - **Desafíos Uruguay^[77]:** Este país divide sus compromisos en doce temáticas:
 - **Participación ciudadana:** Avanzar en la promoción de nuevos espacios de participación ciudadana, así como fortalecer y difundir los espacios ya existentes. Promover actividades en un marco colaborativo que permitan arribar a resultados en beneficio de la ecuación social y en la consolidación de nuevos modelos de relacionamiento con la ciudadanía. Acercar la iniciativa de gobierno abierto a nuevos colectivos sociales.
 - **Monitoreo de Derechos Humanos:** Impulsar iniciativas que transparenten y fortalezcan la procesos nacionales e internacionales; y desarrollar herramientas y promover el acceso a información específica que permita a la comunidad y a grupos de interés desarrollar el monitoreo del cumplimiento de estándares internacionales en materia de derechos humanos.

- **Justicia, anti-corrupción:** Disponibilizar información estadística en materia judicial y abrir canales para encauzar demandas de nueva información; sensibilizar y capacitar sobre temáticas vinculadas al lavado de activos, contribuyendo a mantener y profundizar en la cultura ciudadana la atención sobre estos temas; y generar instancias de discusión con múltiples actores, para impulsar la adecuación normativa relacionada con el lavado de activos.
- **Medio ambiente:** Involucrar a la ciudadanía en el monitoreo de temas relacionados a las políticas medioambientales de la agenda gubernamental y promover el uso y cuidado de los recursos naturales como responsabilidad de toda la sociedad.

Resumen

A lo largo de los años, los gobiernos se vieron con la necesidad de establecer políticas necesarias para la implementación de las iniciativas de datos abiertos y experimentar los beneficios que las mismas conllevan. Gracias a esto lograron no solo mejorar los servicios internos de sus instituciones sino que también alcanzaron niveles de transparencia en cada una de sus áreas, algo que sin dudas generó atención por parte de los ciudadanos. El hecho de que diferentes instituciones regulen y generen estándares y/o métricas para establecer escalas o niveles de apertura de los datos disponibilizados por los gobiernos a nivel mundial, genera un constante compromiso por parte de los mismos y logra, por consiguiente, un mayor interés en mejorar las políticas que corresponden a la transparencia y generación de datos abiertos. Es de esperar que gobiernos que hoy en día no cuentan con portales de datos abiertos o se encuentren en los niveles más bajos en la escala de transparencia, comiencen a tener participación en estos convenios mundiales, organizados por las instituciones mencionadas, con la finalidad de ser parte de un ecosistema mundial que día a día genera mayores oportunidades de desarrollo.

Referencias

- [1] <https://5stardata.info/en/>, fecha de consulta: 11 de marzo 2018.
- [2] <https://www.w3.org/People/Berners-Lee/>, fecha de consulta: 11 de marzo 2018.
- [3] <https://www.w3.org/>, fecha de consulta: 11 de marzo 2018.
- [4] <https://webfoundation.org/>, fecha de consulta: 11 de marzo 2018.
- [5] <https://www.odi.org/>, fecha de consulta: 3 de mayo 2018.
- [6] <https://certificates.theodi.org/en/about/badgelevels>, fecha de consulta: 3 de mayo 2018.
- [7] <http://iso25000.com/index.php/normas-iso-25000/iso-25012>, fecha de consulta: 3 de mayo 2018.
- [8] <http://opendatatoolkit.worldbank.org/es/supply.html>, fecha de consulta: 3 de mayo 2018.
- [9] <http://paquete-apertura-datos.readthedocs.io/es/stable/>, fecha de consulta: 3 de mayo 2018.
- [10] <http://frictionlessdata.io/>, fecha de consulta: 17 de marzo 2018.
- [11] <https://index.okfn.org/>, fecha de consulta: 17 de marzo 2018.
- [12] <https://www.opengovpartnership.org/open-government-declaration>, fecha de consulta: 17 de marzo 2018.
- [13] <http://standard.open-contracting.org/latest/en/>, fecha de consulta: 17 de marzo 2018.
- [14] <http://www.opendatafoundation.org>, fecha de consulta: 18 de marzo 2018.

- [15] <https://www.w3.org/TR/dwbp/>, fecha de consulta: 18 de marzo 2018.
- [16] <https://creativecommons.org/>, fecha de consulta: 19 de marzo 2018.
- [17] <https://datosabiertos.org/en/>, fecha de consulta: 19 de marzo 2018.
- [18] <https://www.idrc.ca/en/initiative/open-data-development>, fecha de consulta: 19 de marzo 2018.
- [19] <https://www.idrc.ca/>, fecha de consulta: 19 de marzo 2018.
- [20] <https://sunlightfoundation.com/>, fecha de consulta: 19 de marzo 2018.
- [21] <https://www.foia.gov/>, fecha de consulta: 19 de marzo 2018.
- [22] <https://okfn.org/network/>, fecha de consulta: 19 de marzo 2018.
- [23] <https://opendatabarometer.org/>, fecha de consulta: 21 de marzo 2018.
- [24] <http://od4d.net/>, fecha de consulta: 21 de marzo 2018.
- [25] <https://www.omidyar.com/>, fecha de consulta: 21 de marzo 2018.
- [26] <https://opendatamonitor.eu/>, fecha de consulta: 22 de marzo 2018.
- [27] https://ec.europa.eu/research/fp7/pdf/fp7-brochure_es.pdf, fecha de consulta: 22 de marzo 2018.
- [28] <https://obamawhitehouse.archives.gov/blog/2013/05/16/introducing-project-open-data>, fecha de consulta: 25 de marzo 2018.
- [29] <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>, fecha de consulta: 25 de marzo 2018.
- [30] <https://www.wikimedia.fr/>, fecha de consulta: 25 de marzo 2018.
- [31] <http://openstreetmap.fr/>, fecha de consulta: 25 de marzo 2018.
- [32] <https://www.etalab.gouv.fr/licence-ouverte-open-licence>, fecha de consulta: 25 de marzo 2018.
- [33] <https://opendatacommons.org/licenses/odbl/>, fecha de consulta: 25 de marzo 2018.
- [34] <https://datos.madrid.es/portal/site/egob/>, fecha de consulta: 25 de marzo 2018.
- [35] <https://www.toronto.ca/city-government/data-research-maps/open-data/open-data-catalogue/>, fecha de consulta: 25 de marzo 2018.
- [36] <https://data.cityofchicago.org/>, fecha de consulta: 25 de marzo 2018.
- [37] <https://opendata.cityofnewyork.us/>, fecha de consulta: 25 de marzo 2018.
- [38] <https://data.london.gov.uk/>, fecha de consulta: 25 de marzo 2018.
- [39] <https://es.wordpress.com/>, fecha de consulta: 26 de marzo 2018.
- [40] <https://datos.gob.mx/>, fecha de consulta: 26 de marzo 2018.
- [41] <https://www.gob.mx/contratacionesabiertas/home#!/>, fecha de consulta: 27 de marzo 2018.
- [42] <http://www.portaltransparencia.gov.br/>, fecha de consulta: 28 de marzo 2018.
- [43] <http://datosabiertos.chilecompra.cl/>, fecha de consulta: 30 de marzo 2018.
- [44] <http://servicios.infoleg.gob.ar/infolegInternet/anexos/265000-269999/265949/norma.htm>, fecha de consulta: 28 de marzo 2018.
- [45] <http://www.buenosaires.gob.ar/aplicacionesmoviles/ba-como-llego>, fecha de consulta: 30 de marzo 2018.
- [46] <http://www.buenosaires.gob.ar/aplicacionesmoviles/ba-ecobici>, fecha de consulta: 30 de marzo 2018.
- [47] <http://www.buenosaires.gob.ar/aplicacionesmoviles/ba-subte>, fecha de consulta: 30 de marzo 2018.
- [48] <http://www.hmabb.gob.ar/habierto/>, fecha de consulta: 30 de marzo 2018.
- [49] <https://www.santafe.gob.ar/ms/gobiernoabierto/colaboracion/santalab/>, fecha de consulta: 4 de mayo 2018.
- [50] <http://www.rosario.gov.ar/web/aplicaciones/aplicaciones-moviles/como-llego>, fecha de consulta: 4 de mayo 2018.
- [51] <http://www.rosario.gov.ar/web/aplicaciones/aplicaciones-moviles/agenda-cultural>, fecha de consulta: 4 de mayo 2018.
- [52] <https://tfl.gov.uk>, fecha de consulta: 25 de marzo 2018.

- [53] <https://api.tfl.gov.uk/>, fecha de consulta: 25 de marzo 2018.
- [54] <https://donde.huesped.org.ar/#/>, fecha de consulta: 25 de marzo 2018.
- [55] <https://github.com/fundhuesped/donde>, fecha de consulta: 25 de marzo 2018.
- [56] <http://www.mejoratuescuela.org/>, fecha de consulta: 25 de marzo 2018.
- [57] <http://atuservicio.uy>, fecha de consulta: 27 de marzo 2018.
- [58] <https://data.issy.com/explore/dataset/disponibilite-parking-aximumcolas/custom/>, fecha de consulta: 27 de marzo 2018.
- [59] <https://www.opendatasoft.es/>, fecha de consulta: 27 de marzo 2018.
- [60] <http://app.wheredoesmymoneygo.org>, fecha de consulta: 27 de marzo 2018.
- [61] <https://www.properati.com.ar/>, fecha de consulta: 5 de mayo 2018.
- [62] <https://www.properati.com.ar/data/>, fecha de consulta: 5 de mayo 2018.
- [63] <http://blog.properati.com.ar/en-que-barrios-las-mujeres-sufren-mas-violencia-en-sus-noviazgos/>, fecha de consulta: 5 de mayo 2018.
- [64] <http://blog.properati.cl/precio-por-manzana-del-sector-oriente/>, fecha de consulta: 5 de mayo 2018.
- [65] <http://blog.properati.cl/ahorrar-viviendo-en-pareja/>, fecha de consulta: 5 de mayo 2018.
- [66] <http://blog.properati.com.mx/2017/11/07/en-que-zona-sismica-esta-tu-vivienda/>, fecha de consulta: 5 de mayo 2018.
- [67] <https://www.properati.com.ar/tools/heatmap>, fecha de consulta: 5 de mayo 2018.
- [68] <http://fedebayle.github.io/potencialesvya/>, fecha de consulta: 5 de mayo 2018.
- [69] <https://www.techo.org/>, fecha de consulta: 5 de mayo 2018.
- [70] https://es.wikipedia.org/wiki/Aprendizaje_autom%C3%A1tico, fecha de consulta: 5 de mayo 2018.
- [71] <https://pulsosocial.com/2016/11/09/ocho-desafios-datos-abiertos-latinoamerica-2017/>, fecha de consulta: 30 de marzo 2018.
- [72] <http://www.labora.io>, fecha de consulta: 13 de abril 2018.
- [73] <http://agacolombia.org/iii-plan-de-accion>, fecha de consulta: 19 de abril 2018.
- [74] https://www.opengovpartnership.org/sites/default/files/Chile_Plan-de-Accion_2016-2018.pdf, fecha de consulta: 19 de abril 2018.
- [75] <https://www.opengovpartnership.org/documents/brazil-national-action-plan-2016-2018-english>, fecha de consulta: 20 de abril 2018.
- [76] <https://www.opengovpartnership.org/documents/argentina-action-plan-2017-2019>, fecha de consulta: 20 de abril 2018.
- [77] <https://www.opengovpartnership.org/documents/uruguay-third-national-action-plan-2016-2018>, fecha de consulta: 21 de abril 2018.

Datos abiertos en universidades

Introducción

Uno de los ámbitos más productivos en la utilización de datos abiertos es el que corresponde a las universidades. La necesidad de generar vínculos entre diferentes comunidades de índole científica o estudiantes de grado y posgrado, y personal docente con el fin de co-gestionar proyectos o mejorar métodos de enseñanza, ha fomentado la iniciativa de datos abiertos en el ecosistema académico. Si bien es un paradigma que recién está en sus comienzos, muchas de las instituciones universitarias que la han adoptado muestran mejoras en avances tecnológicos, transparencia en sus instituciones, colaboración en proyectos de investigación y la innovación de los servicios brindados a sus estudiantes.

En este capítulo se abordan las características y beneficios de los datos abiertos en las universidades. Se detallarán las iniciativas propuestas por varias universidades del mundo y latinoamérica, y se listarán los portales con los que cuentan junto con un breve análisis de los mismos. Además, se mencionarán los casos de éxito de proyectos y aplicaciones con los datos abiertos de las universidades desarrolladas por ellas mismas y por sus estudiantes.

¿En qué consisten?

Ya se ha mencionado el creciente avance de las iniciativas de datos abiertos a nivel mundial y el impacto que estas tienen sobre las partes involucradas. Además, se han expuesto los beneficios que conlleva el uso de datos abiertos por parte de las instituciones, organizaciones y distintas comunidades, fomentando la innovación y creación de nuevos recursos y herramientas.

Las universidades forman parte de un ecosistema gubernamental establecido y cumplen un rol fundamental en la fisonomía sistémica de un gobierno. Son el punto de partida para la generación de nuevos recursos, tanto científicos como de investigación, por lo que no deben quedar a un lado sobre las políticas de datos abiertos. Se debe entender a la universidad abierta como un programa integral que tiene como objetivo incursionar en las políticas de datos y gobierno abierto, y así, lograr un efectivo acceso a la información universitaria y fomentar e incentivar la participación ciudadana en este ámbito.

Cabe destacar la importancia que tiene sentar las bases para que las universidades comiencen a abrir sus datos institucionales, tanto a nivel docente como de investigación, financieros, contractuales, entre otros; adoptando políticas de transparencia conjuntamente con políticas de acceso a la información propuestas por el gobierno involucrado. En este sentido, esta propuesta va más allá de la transparencia y el control en el aspecto administrativo, ya que incursionar en esta iniciativa puede convertir a los datos en la materia prima sobre la cual la sociedad analice y evalúe nuevos esquemas que agreguen valor y fomenten la interacción colaborativa.

Analizando los portales de datos abiertos de universidades del mundo, entre ellos Universidad de Münster (Alemania), Universidad Pablo de Olavide (España), Universidad de Alicante (España), Universidad de Waterloo (Bélgica), Universidad de Exeter (UK), Universidad de Curtin

(Australia) y Universidad de Bologna (Italia) que serán detalladas en la siguiente sección, se puede apreciar que los conjuntos de datos más publicados son:

- **Información organizativa:**
 - Organigramas.
 - Perfiles del personal de la universidad.
 - Autoridades.
 - Ofertas de trabajo en la universidad.
- **Docencia:**
 - Catálogo de estudios.
 - Calendarios y horarios de los cursos.
 - Catálogo de contenidos educativos.
- **Alumnado:**
 - Perfiles de los alumnos matriculados.
 - Programas de becas.
 - Rendimiento académico.
- **Economía:**
 - Estado de cuentas.
 - Presupuesto de ingresos.
 - Presupuesto de gastos.
 - Distribución del presupuesto.
 - Ayudas económicas.
 - Licitaciones.
- **Infraestructura y servicios:**
 - Edificios.
 - Información geoespacial.
 - Información de accesibilidad de los mismos.
 - Equipamiento.
 - Catálogo de servicios informáticos.
 - Servicios a estudiantes.
 - Eventos.
 - Noticias.

Sin embargo, se debe considerar que la iniciativa de universidad abierta o de incorporación de las políticas de datos abiertos a las universidades, es muy reciente y a su vez conlleva un arduo trabajo que debería estar en la agenda de los diferentes gobiernos y responsables de dichas instituciones para crecer y mejorar paulatinamente.

¿Cuáles son sus beneficios?

La apertura de datos de las universidades es el punto de partida para generar comunicación y conexión entre diferentes áreas de investigación conjunta que engloban a las diferentes universidades locales. Además, esta iniciativa permite establecer transparencia en las instituciones, detectar oportunidades de mejora en los servicios académicos y generar una mayor participación activa por parte del alumnado como consecuencia del aspecto colaborativo.

Entre los diferentes datos que se pueden encontrar en el ambiente universitario, los que se consideran de mayor interés son aquellos que están ligados a actividades de investigación, ya que son una de las dos actividades principales que la universidad brinda como servicio a la sociedad; la

otra es la educación en sí misma. La apertura de este tipo de datos tiene como beneficio el conocimiento constante sobre la financiación de las investigaciones y la generación de vínculos entre investigadores, entre sí y con potenciales entidades interesadas en aplicar o apoyar su trabajo.

Entre los beneficiarios de la apertura de datos planteada, no solo se encuentra la propia universidad, sino también los diferentes actores de la administración educativa como padres, maestros, estudiantes, investigadores, ciudadanos, gobierno adjunto, entre otros. El acceso abierto a los datos administrativos académicos debería permitir mejorar la eficiencia de dicha gestión y la toma de decisiones a partir de un análisis exhaustivo de la información. Además, se logra un vínculo estrecho entre quienes monitorean el gasto educativo, su gestión y su eficiencia, y permite la rendición de cuentas establecida en las políticas de gobierno abierto.

Otros de los aspectos más relevantes de los beneficios de la apertura de información universitaria son:

- **Su aporte intrínseco al desarrollo educativo:** los datos abiertos permiten a los estudiantes aprender a resolver problemas reales, desarrollar conocimiento y fomentar el vínculo con la investigación colaborativa. Además, establecen nuevos formatos de contenidos y orientan al alumnado hacia el aprendizaje de nuevas herramientas para el análisis, visualización y procesamiento de los datos.
- **Su valor científico:** los datos abiertos son el principal fuerte de la Ciencia Abierta, ya que facilitan no sólo la investigación científica, sino que además, promueven su transparencia, facilitando a otros investigadores la utilización de resultados y procesos llevados adelante, así como también los datos crudos para combinarlos con otros, expandiendo los límites del desarrollo del conocimiento.
- **Sus aportes a la transparencia y la rendición de cuentas:** las universidades públicas deben rendir cuentas como cualquier otra institución pública con el fin de que la comunidad pueda comprender cómo se ejecutan los fondos, quiénes y cómo son contratados, cómo es la distribución de cargos y las escalas salariales.

Temas como la desigualdad de género en los cargos jerárquicos académicos o sesgos en los índices de graduación, así como aspectos sobre los gastos o distribución de fondos entre las diferentes carreras pueden ser puestos en valor para generar conciencia sobre datos concretos y agilizar una mejor toma de decisión. Además, la utilización de datos como fuente permite hacer frente con argumentos concisos a casos de corrupción de los cuáles las instituciones educativas no están exentas. En resumen, permite colaborar de forma conjunta, entre los distintos actores del ámbito académico, para que las universidades cumplan su rol social y político.

Investigación abierta, acceso abierto y comunidad colaborativa

Es importante destacar tres conceptos importantes que se desprenden del paradigma de apertura de la información, y hacerlo en este capítulo correspondiente al ámbito académico no es casualidad ya que éste es el precursor y principal fuente de investigación colaborativa mediante diferentes programas de aprendizaje; los mismos son: investigación abierta, acceso abierto y comunidad colaborativa.

Para empezar, se mencionarán paralelismos entre dos culturas que fomentan la apertura de la información, el acceso a ella y a los procedimientos para su obtención y, a su vez, expanden la participación conjunta para lograr nuevos objetivos: la cultura científica y la cultura hacker. No sería extraño trazar un paralelismo entre estas dos culturas, ya que la que corresponde a los hacker tiene en sus inicios estudiantes universitarios fuertemente impregnados en la cultura académica de intercambio libre de la información y conocimientos, propias de toda ciencia abierta.

Se puede mencionar en primer lugar, entre las similitudes, el carácter meritocrático de ambas culturas. En segundo lugar, el carácter colectivo de la creación; si bien se valora y reconocen las intervenciones individuales, los resultados son un bien de la comunidad. En tercer lugar, los resultados se comunican de forma abierta y sin restricciones, mediante diferentes mecanismos de difusión. Por último, dichos resultados son evaluados por la comunidad y, de su evaluación depende el éxito o el fracaso de la creación.

A su vez, presentan diferencias importantes, entre las que se encuentra el reconocimiento certificado que persigue la cultura científica; los miembros de la comunidad científica aspiran y compiten por la obtención de acreditaciones oficiales que les permitan avanzar en su carrera profesional. Por otra parte, la comunidad hacker no solo hace público el resultado, sino también el proceso mediante el cual se obtuvo dicho resultado. Por último, el sistema de revisión por pares es abierto y además se extiende a la post-publicación.

Diversas iniciativas y movimientos se han puesto en marcha bajo los conceptos antes mencionados ya que lo que se conocía como ciencia abierta en sus inicios se encontraba limitada a ciertas observaciones subjetivas, ya que muchos de los resultados científicos pueden terminar siendo apropiados y posteriormente comercializados, por lo cual se intenta fomentar la desvinculación de este tipo de actividades mediante la incorporación de los conceptos específicos de la actividad participativa.

Uno de los primeros intentos fueron las publicaciones de revistas científicas, pero si bien hoy en día se pueden encontrar en su versión digital, se pueden acceder mediante suscripciones o pedidos a bases de datos a costos elevados. A partir de 2008, varias universidades adoptaron políticas de promoción del acceso abierto a las publicaciones de profesores e investigadores mediante repositorios institucionales, entre ellas las correspondientes a la Unión Europea. Además, quien se unió a la iniciativa fue la Agencia Nacional Institutes of Health, del Departamento de Salud norteamericano ya que figuraba como la segunda institución más importante en lo que a financiamiento de investigación corresponde.

Un concepto muy importante en este aspecto es el de acceso abierto^[1] (OA, Open Access) ya que persigue el objetivo de permitir a los consumidores de información académica tener acceso a ella sin costos elevados. Esto puede ser posible mediante licencias del tipo Creative Commons, pero tienen la principal limitación de que las revistas científicas podrán atender este pedido recién al año de ser publicadas.

El otro concepto importante antes mencionado es el de investigación abierta^[1] (OR, Open Research), que persigue la apertura del proceso de la obtención del resultado científico luego publicado. Por lo tanto, no solo cuestiona el acceso parcial a los resultados finales, sino que plantea el acceso a todos los elementos potencialmente útiles para futuros investigadores.

Por último, introducir el concepto de comunidad colaborativa a la cultura científica que en sí es una actividad colaborativa por excelencia no tiene mucho sentido, por lo que se aclara que el concepto se refiere a los mecanismos de colaboración. En los comienzos, estos mecanismos

resultaban tardíos o carentes de tecnología. En los tiempos que corren, el avance tecnológico fue de gran aporte para agilizar estas comunicaciones y resoluciones metódicas de los investigadores involucrados. Tan es así, que hace algunos años se generó lo que se conoce como ciencia ciudadana^[1] (CS, Citizen Science), que permite la participación de ciudadanos no-expertos en proyectos de ciencia estándar. Una de las primeras experiencias al respecto fue Nasa Clickworkers^[2], un proyecto piloto propuesto por la Nasa para catalogar cráteres en Marte, a partir de las imágenes obtenidas por la sonda Viking. Una experiencia más reciente es Galaxy Zoo^[3], un proyecto de astronomía que utiliza voluntarios para clasificar millones de galaxias, según las imágenes tomadas por el Sloan Digital Sky Survey. Este proyecto es producto de la colaboración entre la Universidad de Oxford, la Universidad de Portsmouth, la Universidad Johns Hopkins y la Universidad de Yale.

Iniciativa de datos abiertos universitarios en el mundo

A partir del reciente crecimiento de los datos abiertos en universidades, varios gobiernos adoptaron una postura de avance hacia una gestión y administración transparente sobre este tipo de instituciones, con el fin de instalar las políticas necesarias para mejorar la eficiencia de sus áreas.

El gobierno es el principal actor en este aspecto, ya que es quien establece las políticas de acceso a la información que serán la base para que las distintas instituciones académicas puedan recabar información, analizarla y ponerla a disposición de la comunidad bajo los lineamientos de datos abiertos vistos en secciones anteriores. No solo son importantes la política y las leyes abordadas por el gobierno, sino también el disponer de una estructura válida y consensuada entre las diferentes áreas de educación para determinar qué datos serán presentados de manera abierta y cómo serán disponibilizados, dependiendo de las herramientas tecnológicas.

En esta línea se mueven gobiernos como los de Gran Bretaña y España, cuyas organizaciones educativas establecen los parámetros y planes de acción sobre los cuales las distintas universidades se basan para lograr una apertura de datos de calidad. Entre los parámetros y planes de acción mencionados se encuentran definidas las condiciones que deciden qué datos deben ser disponibilizados y cuáles no, haciendo foco no solo en aspectos institucionales, sino también edilicios, financieros y de índole educativa. Todos los portales de datos abiertos de universidades cuentan con un alto nivel en la escala de calidad de apertura mencionada en el capítulo anterior. Esto permite a la comunidad de datos abiertos proporcionar nuevas herramientas y aplicaciones a partir de datos académicos y, a su vez, disponer de visualizaciones interactivas que resulten amigables al consumidor de la información.

Entre las universidades que han logrado un gran avance en las políticas de datos abiertos se encuentran:

La iniciativa de datos abiertos de la Universidad de Münster^[4] (Alemania) tiene como punto de partida la interacción de las corrientes de investigación de la institución y la creciente colaboración entre los diversos investigadores y estudiantes avanzados con el fin de lograr la creación de una comunidad que co-gestione proyectos científicos. La iniciativa de esta universidad adopta el nombre de LODUM^[5] (Linked Open Data University of Münster) y comenzó con financiamiento exclusivamente interno hasta que logró la colaboración de organismos externos para generar los recursos económicos necesarios. Esta universidad es la primera que incursiona en la política de datos abiertos en Alemania y establece los parámetros a seguir correspondiendo con los lineamientos propuestos por el Consejo de Investigación Alemana. Este plan de apertura y transparencia de la información persigue el objetivo de centralizar la información de las facultades

que corresponden a la universidad y, por consiguiente, ampliar el radio de usuarios que se beneficien con dicha arquitectura, permitir conexiones flexibles para nuevos emprendimientos y extender las temáticas de la información.

Es importante destacar el proyecto que lleva adelante la Fundación de Investigación Alemana^[6] (DFG, por sus siglas en alemán) en conjunto con las diversas instituciones académicas de la Universidad de Münster tales como el Laboratorio de Interoperabilidad Semántica de Münster^[7], el Instituto de Geoinformática^[8] y la Biblioteca de la Universidad^[9], que se conoce como LIFE^[10] (Linked Data for eScience Services) y tiene como objetivo mejorar la colaboración en investigación y educación por medio del intercambio de los datos científicos organizados por tiempo, espacio y semántica. Mediante este proyecto, la universidad pretende lograr mayor fluidez en la comunicación de los proyectos de investigación de sus estudiantes y usuarios ligados a los mismos de formas alternativas, y complementar la iniciativa LODUM mencionada anteriormente, ya que los nuevos colaboradores de las investigaciones científicas en curso podrán acceder de forma remota y sin restricción alguna al repositorio centralizado de la información.

La iniciativa de datos abiertos de la Universidad Pablo de Olavide de Sevilla^[11] (UPO) tiene su origen en las II Jornadas de Gobierno Abierto y Opendata Sevilla^[12], en las que se presentó una versión beta del portal de datos abiertos de la institución. Desde ese momento, la UPO se ha comprometido con las iniciativas de gobierno abierto llevando a cabo experiencias de procesos participativos, en las cuales se logró, por ejemplo, elaborar el Plan Estratégico de la Universidad^[13] que consiste en el proceso de mejora de aspectos relevantes de la institución como el compromiso social, el compromiso institucional, la innovación continua, los objetivos académicos de los estudiantes, la capacitación del personal docente, la reestructuración y eficacia de la administración institucional, entre otros aspectos. En 2014, se presentó la versión estable del portal de datos abiertos utilizando CKAN como framework para la gestión de los datos. En dicha plataforma (<https://datos.upo.gob.es/>) los datos se encuentran publicados predominantemente en formato csv y son disponibilizados directamente desde el sistema de información de la universidad. Además, se presentan diferentes visualizaciones de los conjuntos de datos publicados a través de herramientas alternativas que optimizan su procesamiento, como OpenSpending^[14] y Tableau Public^[15].

Es importante destacar su incorporación al grupo de trabajo de la Open Data España, la comunidad que reúne a las administraciones públicas españolas, que trabajan en el avance de las políticas de apertura de datos; y a entidades y profesionales expertos en la materia, como la Fundación CTIC^[16], socio español de W3C.

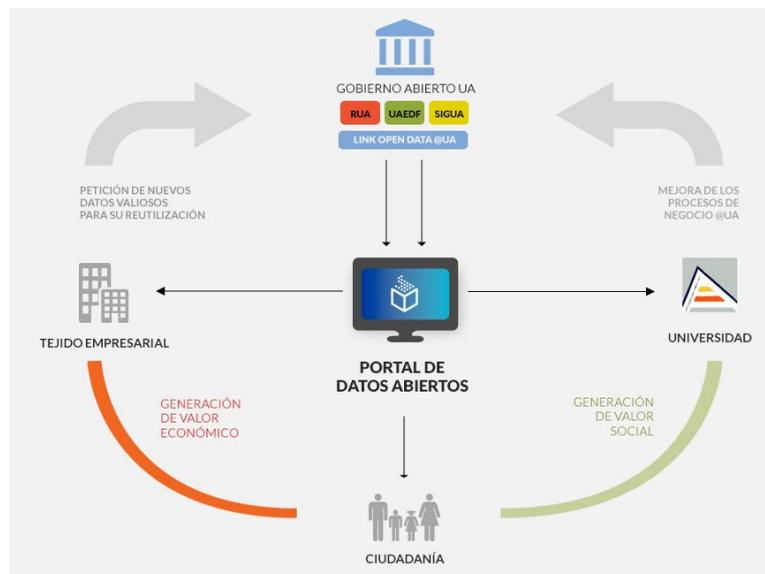
El portal de datos abiertos de la universidad presenta conjuntos de datos agrupados en categorías como educación, sector público y energía. Dentro de las mismas aparecen datos interesantes como presupuestos y consumos realizados en el ámbito académico. Además, contiene un canal de interacción para poder reunir ideas y sugerencias que permitan mejorar la comunicación de la información basado en la experiencia de los usuarios. Por último, dispone de una sección de noticias en la que se muestran casos de evolución de los presupuestos estimados y actividades universitarias de participación ciudadana y colaboración como hackatones.

La iniciativa de gobierno abierto de la Universidad de Alicante^[17] obtuvo un valor fundamental dentro de las políticas de IT en su nuevo plan estratégico 2014/2019^[18]. Una de las ideas prioritarias de esta iniciativa es el desarrollo de un proyecto de reutilización de la Información de la Universidad de Alicante conocido como OpenData4U^[19].

El principal objetivo que persigue la institución a través del proyecto OpenData4U es crear metodologías estables para instalar una política de datos abiertos clara en su esquema

organizacional. Esto da la oportunidad de fomentar y ejercitar el desarrollo de nuevos conocimientos y vínculos entre estudiantes avanzados y egresados para facilitar su inserción laboral y sentar las bases de una comunidad sostenible mediante la creación de servicios digitales.

Uno de los primeros frutos de esta iniciativa fue la puesta en marcha del Portal de Datos Abiertos de la Universidad (<https://datos.ua.es/>) que presenta un nutrido abanico de conjuntos de datos correspondientes a áreas como investigación, economía, becas y actividades, y personal, entre otros. Además, cuenta con una sección específica para aplicaciones elaboradas por terceros con los datos abiertos disponibles que serán abordadas en secciones posteriores.



Interacción y actores de la políticas de datos abiertos de la Universidad de Alicante.

La Universidad de Alicante impulsa desde hace tiempo iniciativas relacionadas con el gobierno abierto que tienen como principal objetivo potenciar la transparencia y las acciones colaborativas.

Existen varios proyectos IT en la universidad que están relacionados con la plataforma de datos abiertos como Universidad en Cifras^[20], que muestra el análisis de diversos indicadores de la mejora continua de la calidad en el ámbito universitario. Otro ejemplo es el Sistema de Información Geográfica de la Universidad de Alicante (SIGUA)^[21], que muestra un mapa con las ubicaciones de las distintas áreas correspondientes a la universidad, ya sean facultades, centros de actividades, centros comerciales o zonas recreativas. Por otra parte, se encuentra el Portal de la Educación Digital del Futuro (UAedf)^[22] que nuclea todo el plan docente incluyendo herramientas, cursos, material académico, programas y metodologías. Por último, el Repositorio Institucional de la Universidad de Alicante (RUA)^[23] que ofrece de forma abierta documentos docentes y de investigación generados por los miembros de la universidad.

Otros ejemplos de iniciativas de datos abiertos en universidades del mundo son: el portal beta (<https://dati.unibo.it/>) de la Universidad de Bologna de Italia^[24] que se encuentra basado en CKAN y disponibiliza información detallada sobre los cursos, planes de estudio, personal e informes sociales de la universidad; el repositorio de investigación (<https://ore.exeter.ac.uk/repository/>) de la Universidad de Exeter, Gran Bretaña^[25], que contiene datos en formato abierto para generar una comunicación directa con investigadores de las distintas temáticas; la plataforma de investigaciones (<https://researchdata.and.s.org.au/>) de la Universidad de Curtin, Australia^[26] donde se pueden realizar búsquedas para conocer los proyectos en los que trabajan miembros del establecimiento y utilizarlos

para aportar nuevos hallazgos; el Github^[27] (<https://github.com/uWaterloo/Datasets>) de la Universidad de Waterloo^[28], Bélgica, donde se pueden encontrar conjuntos de datos relacionados al eje administrativo como salarios, cursos, vacaciones del personal y pagos del estacionamiento de la institución, y la interfaz de acceso (<https://uwaterloo.ca/api/>), que pone a disposición la universidad, a través de la cual se puede acceder programáticamente a la información antes mencionada.

Es de esperar que las nuevas corrientes de datos abiertos en universidades a nivel mundial tomen mayor relevancia con el tiempo, como así lo ha hecho la iniciativa de gobierno abierto. De todas maneras, esto se ve directamente relacionado con la toma de decisiones políticas de cada gobierno, salvo casos excepcionales. En la medida que los mismos incursionen en los beneficios que produce llevar este paradigma al ámbito académico, las instituciones mundiales se verán conectadas en una gran red de información colaborativa.

Iniciativa de datos abiertos universitarios en Latinoamérica y Argentina

Existe una brecha muy grande entre el avanzado estado en que se encuentran las universidades de España y Reino Unido, que cuentan con niveles 4 o 5 en la escala de estrellas de apertura de la información presentada en el capítulo anterior, y las universidades regionales que incursionan en el concepto de transparencia y que no ascienden del primer nivel en la escala mencionada, sumando la escasa cantidad de establecimientos que tienen esta temática en agenda.

Estas universidades presentan gran interés en confeccionar su plan estratégico de transparencia poniendo a disposición en sus plataformas información de gran valor como presupuestos, normativas y datos institucionales y organizativos, sin embargo, evitan entrar en detalle. La mayoría de las universidades regionales no cuentan con un portal de datos abiertos propiamente dicho, es decir con los datos disponibilizados en formato reutilizable y con canales de comunicación directa con el ciudadano. Esto determina que las iniciativas de datos abiertos en las universidades de Latinoamérica recién están ingresando en la consideración de los gobiernos.

La Universidad Nacional Autónoma de México^[29] es una de las escasas instituciones que pone a disposición un Portal de Datos Abiertos^[30]. En él se encuentran colecciones de investigación académica de diferentes comunidades en formato descriptivo y descargable. A su vez, presenta diferentes gráficos a modo de representación de la información con la que cuenta el portal para dar un panorama de las temáticas involucradas.

Por otra parte, en Brasil, se encuentran casos como la Universidad de San Pablo^[31] o la Universidad de Brasilia^[32] que ponen a disposición información detallada de sus finanzas y procesos licitatorios. Todos los datos disponibilizados por estas instituciones se encuentran en formatos pdf y en diferentes plataformas debido a que no cuentan con un portal central que nucleee la información.

Por ejemplo, la Universidad de Brasilia cuenta con diferentes archivos que dan valor sobre sus obras de construcción en su web de transparencia, aunque en formatos no estructurados.

The screenshot shows the website of Universidade de Brasília. At the top, there is a navigation bar with links for 'UnB Webmail', 'Portal Aluno', 'CPD', 'Telefones', 'Sistemas', 'MençãoWeb', 'SaeWeb', and a search box. Below this is the university's logo and name. A secondary navigation bar includes links for 'página inicial', 'noticias', 'estude na UnB', 'aluno de graduação', 'aluno de pós-graduação', 'ex-aluno', and 'servidor'. A large banner image shows a group of people walking on a path. To the left, there are several menu categories: 'UnB para a comunidade', 'Serviços no campus', and 'Telefones da UnB'. The main content area is titled 'Editais para obras de construção' and lists several notices, including 'RESERVATÓRIO - RDC 002/2017', 'LDTEA - RCD 001/2017', and 'Engenharia Florestal - CC 134/2016'.

Datos no estructurados de las obras de construcción de la Universidad de Brasilia.

Por su parte, la Universidad de San Pablo cuenta con la misma consideración sobre los datos disponibilizados, ya que la mayoría se encuentra en formato no estructurado o en enlaces vinculados a plataformas alternativas de búsqueda dentro de los propios sitios de la universidad y cuyos resultados son listados estáticos de información. Tal es el caso de los convenios que tiene la universidad con las diversas entidades y que corresponden al área de administración y finanzas.

The screenshot shows the website of USP (Universidade de São Paulo). It features the university's logo and name. Below the header, there is a search bar with the text 'Pesquisar Convênios'. The search results are displayed in a table with the following columns: Documento, Situação, Unidade, Processo, Modalidade, Submodalidade, Objeto, Coordenador, and Vice. The table contains 18 rows of data, with the 12th row highlighted in yellow.

Documento	Situação	Unidade	Processo	Modalidade	Submodalidade	Objeto	Coordenador	Vice
1944	Aprovado	4 - IEE	00.1.00057.04.7	Convênio	Convênio	Regular as condições e respon	2261983-Murilo Tadeu Wierne	-
15863	Aprovado	4 - IEE	06.1.00372.04.5	Convênio	Convênio	Eliminar a dependência de eq	2310754-Roberto Zilles	-
13607	Aprovado	4 - IEE	06.1.00582.04.0	USP Contratada	USP Contratada	CONTRATO FIRMADO ENTRE F	59110-Jose Aquiles Beasso Gr	-
13815	Para Regularizar	4 - IEE	07.1.00147.04.2	USP Contratada	USP Contratada	Prestação de serviços de cons	59110-Jose Aquiles Beasso Gr	-
20130	Aprovado	4 - IEE	09.1.00346.04.7	Auxílio Individual	Auxílio Individual	Execução de missões de traba	502232-Elvo Callisto Burini Jur	-
17933	Para Regularizar	4 - IEE	06.1.00397.04.8	Convênio	Convênio	O presente termo tem por obj	59110-Jose Aquiles Beasso Gr	-
21343	Aprovado	4 - IEE	09.1.00664.04.9	Pré-Projeto	Pré-Projeto	O presente projeto estabeleci	59110-Jose Aquiles Beasso Gr	-
21479	Aprovado	4 - IEE		Convênio	Convênio	A Concedente poderá concede	59110-Jose Aquiles Beasso Gr	-
21480	Aprovado	4 - IEE		Convênio	Convênio	Estágio Estudantes	59110-Jose Aquiles Beasso Gr	-
21738	Aprovado	4 - IEE	10.1.00081.04.5	Protocolo Intenção	Internacional	Desenvolvimento de projectos	59110-Jose Aquiles Beasso Gr	-
21215	Aprovado	4 - IEE	09.1.00623.04.0	Convênio	Convênio	Consolidação da Infra-Estrut	52176-Alexandre Plantini	-
14674	Para Regularizar	4 - IEE	07.1.00461.04.9	Convênio	Convênio	Convênio entre a Petrobrás,	1889737-Suani Teixeira Coelh	-
21519	Aprovado	4 - IEE	10.1.00013.04.0	Convênio	Acadêmico Internacional	Promover a cooperação acad	2564153-Neli Aparecida de M	-
21735	Aprovado	4 - IEE	10.1.00082.04.1	Protocolo Intenção	Internacional	Desenvolvimento de pesquisa,	59110-Jose Aquiles Beasso Gr	-
15162	Para Regularizar	4 - IEE	07.1.00609.04.6	Convênio	Convênio	Levantamento Georreferenci	1889737-Suani Teixeira Coelh	-
21517	Para Regularizar	4 - IEE	09.1.00672.04.1	Convênio	Convênio	Estágio Estudantes	59110-Jose Aquiles Beasso Gr	-
21516	Para Regularizar	4 - IEE	09.1.00671.04.5	Convênio	Convênio	Estágio Estudantes	59110-Jose Aquiles Beasso Gr	-
21542	Para Regularizar	4 - IEE	10.1.00032.04.4	Convênio	Convênio	Oferecimento de estágios de e	59110-Jose Aquiles Beasso Gr	-

Resultado no estructurado de los convenios de la Universidad de San Pablo.

Chile, por su parte, con la Universidad de Atacama^[33] o la Universidad de Chile^[34], cuenta con el mismo esquema que las universidades brasileñas antes mencionadas. Es decir, dispone de una sección destinada a la transparencia dentro de los sitios institucionales de las universidades, que engloba información relevante sobre compras, licitaciones, trámites, actividades, presupuestos y beneficios; pero la presenta en forma de índice con enlaces que dirigen a distintas secciones donde se encuentra realmente la información concreta del tema abordado en formato no estructurado.

Sección de transparencia de la Universidad de Atacama (Chile)

Sección de transparencia de la Universidad de Chile

Por otra parte, existe un caso alentador en la región que es el de la Universidad Central del Ecuador^[35]. Esta institución cuenta con un Portal de Datos Abiertos (<http://datosabiertos.uce.edu.ec/>) que actúa como plataforma única de concentración de información en formato estructurado (.xls). Si bien la información disponible hasta el momento es escasa, ya que cuenta únicamente con datos de matriculados, la misma permite la generación de indicadores a modo de gráficos que facilitan un muestreo de la cantidad de matriculados por facultad o especialidad y un panorama de la forma de contratación de los docentes de la universidad.



Gráficos sobre los matriculados de la Universidad Central del Ecuador.



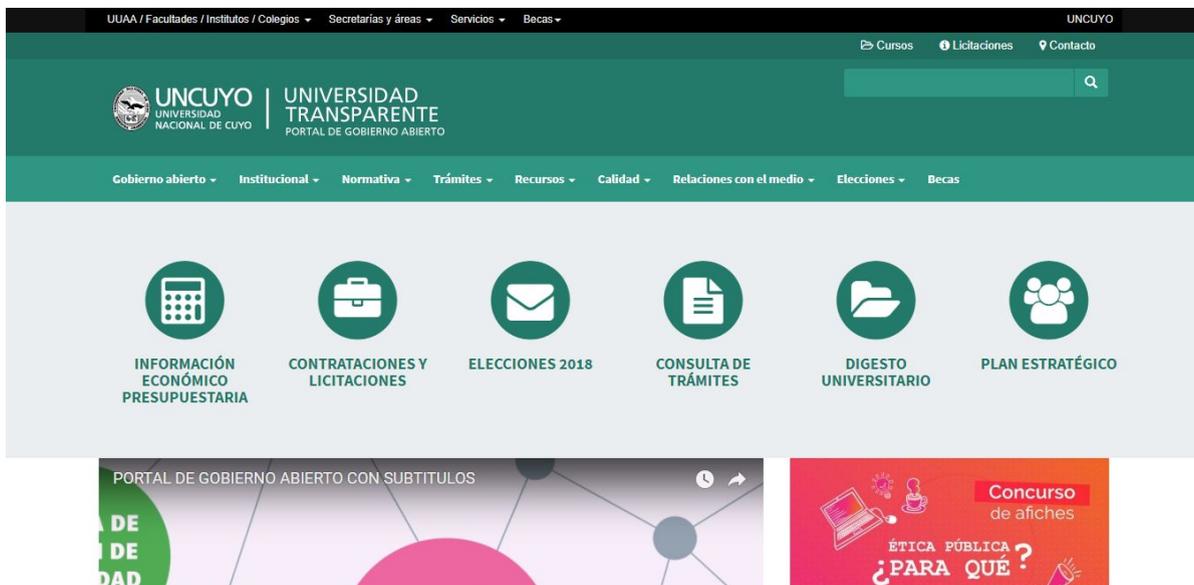
Gráficos sobre los docentes de la Universidad Central del Ecuador.

En lo que respecta a las iniciativas de datos abiertos en el ámbito universitario de Argentina, el panorama es el mismo que en la mayoría de la región. Los datos abiertos en el ámbito universitario a nivel nacional son escasos o prácticamente nulos. La disponibilidad de la información, el acceso a la misma y las buenas prácticas mencionadas en capítulos anteriores, no se encuentran visibles en el ambiente académico, al menos en el sector público.

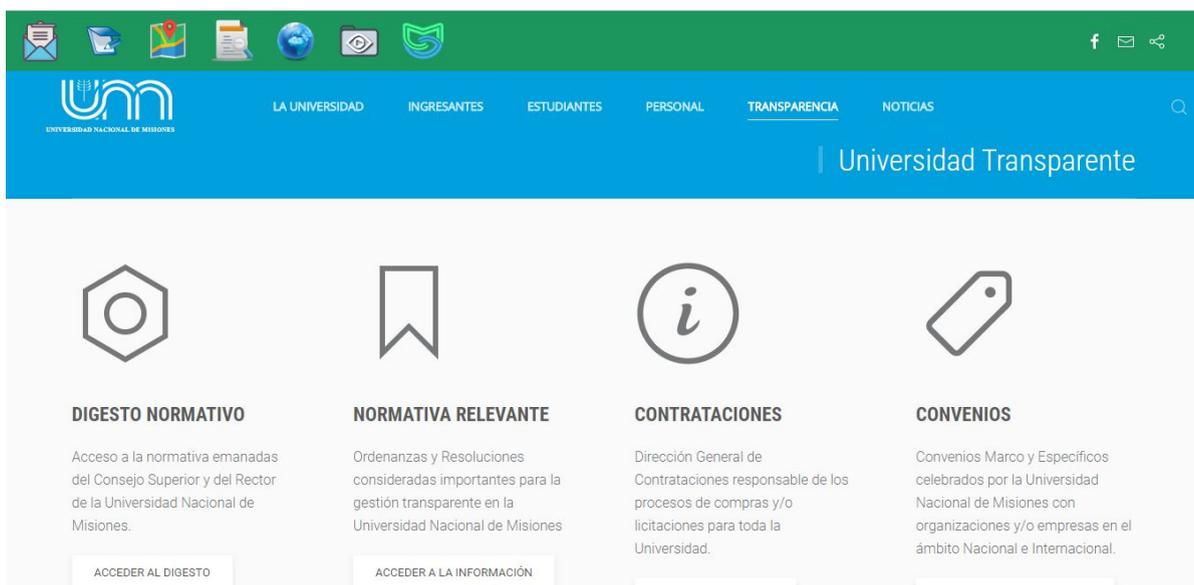
Algunos casos como la Universidad Nacional de Cuyo (UNCUYO)^[36] y la Universidad Nacional de Misiones (UNAM)^[37] afirman esto. Los datos disponibilizados por las entidades se encuentran en el primer nivel de la escala de apertura de datos debido a que no disponen de portales o plataformas donde se centralice la información, no se efectúa la apertura en formatos reutilizables ni se fomenta la interacción con el ciudadano sobre estos conceptos.

Los portales de las universidades mencionadas cuentan con un gran repositorio de datos basados en resoluciones y ordenanzas; y contrataciones y presupuestos, entre otras cosas. Esta

información se presenta en una sección específica de transparencia dentro del sitio institucional. Es importante destacar que se muestra en detalle el proceso licitatorio de las contrataciones realizadas pero en formato no estructurado.



Sección de transparencia de la Universidad Nacional de Cuyo (Argentina)^[38]



Sección de transparencia de la Universidad Nacional de Misiones (Argentina)^[39]

Uno de los sistemas argentinos que vale la pena destacar es el Sistema de consulta de estadísticas universitarias^[40]. Esta plataforma es un desarrollo conjunto entre la Secretaría de Políticas Universitarias y la Universidad de Cuyo, y permite realizar consultas sobre el conjunto de información de todas las instituciones universitarias, ya sean privadas como públicas, que conforman el amplio ecosistema académico. Este sistema brinda de manera interactiva una herramienta que permite establecer una conexión entre los estudiantes de pregrado, grado y postgrado, y las condiciones académicas. Se puede visualizar la cantidad de estudiantes respecto de la carrera universitaria que cursa, hacer distinciones entre géneros, condición de la institución (privada o pública) y condición del personal docente de la institución (titular, adjunto, jtp), entre los gráficos más destacados. La principal limitación de este sistema es que no se encuentran disponibles los datos

con los que se generan dichos reportes y estadísticas, lo que podría ser de gran utilidad para realizar nuevos proyectos.

Ante el análisis sobre las universidades latinoamericanas, es importante destacar la iniciativa que abordan cada una de las instituciones apoyadas en las políticas establecidas por sus gobiernos para lograr un mayor grado de transparencia en cada una de sus áreas y gestionar un crecimiento sostenible de las comunidades de datos abiertos que, día a día, trabajan en la generación de mejoras y nuevas herramientas para el beneficio social.

Casos de éxito

Es sabido que el concepto de datos abiertos en el ámbito universitario está teniendo un paulatino crecimiento y que sus logros permiten no solo incursionar en escalas notorias de transparencia, sino también abastecer de materia prima a las comunidades encargadas de explotar la información para gestionar y co-crear nuevos proyectos y corrientes de investigación. Como se mencionó en las secciones anteriores, los estudiantes se ven incluidos en proyectos colaborativos que tienen un impacto resonante en la toma de futuras decisiones, no solo a nivel universitario sino también a nivel gubernamental, ya que dicho producto puede tomar grandes dimensiones de interés general. A su vez, el personal docente establece nuevos paradigmas en sus metodologías de enseñanza y se nutre de las herramientas que proveen las áreas ligadas al desarrollo tecnológico utilizando los datos abiertos disponibilizados por la institución en cuestión.

A continuación se mencionan los proyectos con mayor impacto en la vida cotidiana de los estudiantes y docentes que integran el ecosistema académico y que fomentan la réplica en diferentes instituciones universitarias del mundo a partir de los datos abiertos generados por las mismas.

- **College Scorecard^[41]**: Es un proyecto propuesto por el Departamento de Educación de Estados Unidos^[42] que permite a los futuros estudiantes o estudiantes regulares buscar y comparar universidades dependiendo de los programas que disponen, duración de la carrera, ubicación y tamaño de la institución, características especiales de la universidad, entre otras. Sus resultados permiten hacer estimaciones de financiamiento de la universidad y promedio de graduados en dicha institución. Además, el proyecto cuenta con una plataforma con los datos abiertos^[43] que genera el buscador propiamente dicho. Estos datos son disponibilizados a través de una API^[44] para ser consumida por terceros y posibilitar la generación de nuevos proyectos.
- **Campusplan^[45]**: Es una aplicación de la Universidad de Münster de Alemania que utiliza los datos abiertos dispuestos por la institución a través de su portal de iniciativa de datos abiertos LODUM. Esta aplicación permite al estudiante acceder a información detallada de las localizaciones de los recursos edilicios correspondientes a la universidad, horarios de los cursos, agenda de las habitaciones que corresponden al campus y menús que presenta el comedor interno de la institución.
- **UAlojamientos^[46]**: Es una aplicación para teléfonos móviles desarrollada por el Servicio de Informática de la Universidad de Alicante, que reutiliza los datos de alojamiento y alquileres de pisos de la Oficina de Información al Alumno de la Universidad de Alicante para ayudar a los estudiantes en la búsqueda del mejor alojamiento.

- **Science Area Map^[47]**: Es un mapa generado por la Universidad de Oxford que muestra, a partir de los datos abiertos de la universidad y los puntos geométricos con los que cuentan, las diversas áreas de ciencia de la universidad que comprende el campus. Entre ellos se encuentran los edificios de investigación, aulas, departamentos administrativos, oficinas de salud y seguridad, entre otros.
- **Room finder^[48]**: Es un sistema de búsqueda de salas de la universidad que el alumno puede necesitar para diversos propósitos como reuniones, charlas y conferencias. Este software aglomera los diferentes datos abiertos de la universidad para hacer las comparaciones y filtros necesarios en dicha búsqueda; entre los datos se encuentran las características de la sala y los horarios de uso, la disponibilidad de los edificios de la institución y el conjunto de datos de reservas propiamente dicho.
- **WatMine^[49]**: Es una extensión para el navegador web Chrome de Google desarrollada por estudiantes de la Universidad de Waterloo de Bélgica, que permite asociar el usuario de la universidad con dicha extensión para poder interactuar con los sitios correspondientes al ecosistema de la universidad de forma rápida y efectiva. Establece vínculos entre los artículos a los cuales el usuario accede y los cursos a los que concurre el estudiante. Además, cuenta con un sistema de notificaciones al usuario del navegador web sobre el estado de cuenta de los beneficios que provee la universidad al estudiante en cuestión.

Resumen

Las necesidades de colaboración por parte de las diferentes áreas de investigación científica y, por consiguiente, la mejora sustancial en la reutilización de los datos obtenidos por cada una de ellas para agilizar y expandir el ecosistema de proyectos de dicha índole, generó el concepto de datos abiertos. En consecuencia, se puede decir que es el ámbito académico el impulsor de la importancia de los datos abiertos en el mundo.

El auge de las iniciativas de gobierno abierto a nivel mundial, llevaron a que los gobiernos tomen consideración en la implementación de este paradigma en las instituciones académicas con el fin de mejorar el nivel de transparencia en ellas y fomentar el vínculo con los estudiantes mediante conferencias o eventos abordando dicha temática y lograr así un incremento en la comunidad de datos abiertos. Además, la apertura de datos colabora en el ámbito universitario brindando a los estudiantes, investigadores y docentes, información confiable para trabajar, analizar, estudiar y generar nuevos proyectos y metodologías de enseñanza.

Si bien este concepto va ganando terreno en este tipo de instituciones, es prácticamente nuevo y es por este motivo que los países de latinoamérica se encuentran relegados respecto a sus pares en Europa o América del Norte. No obstante, hay un gran desafío propuesto por cada uno de ellos para inculcar en sus instituciones académicas esta iniciativa.

Referencias

[1] Gonzalez García, Marta y Lopez Cerezo, J. (2012) *Fronteras de la Ciencia. Hibridaciones*. Ed. Organización de los Estados Iberoamericanos, fecha de consulta: 31 de julio 2018.

- [2] <http://www.nasaclickworkers.com/>, fecha de consulta: 31 de Julio 2018.
- [3] <https://www.zooniverse.org/projects/zookeeper/galaxy-zoo/>, fecha de consulta: 31 de Julio 2018.
- [4] <https://www.uni-muenster.de>, fecha de consulta: 30 de mayo 2018.
- [5] <https://lodum.de>, fecha de consulta: 30 de mayo 2018.
- [6] <http://www.dfg.de>, fecha de consulta: 30 de mayo 2018.
- [7] <https://www.uni-muenster.de/Geoinformatics/en/musil/index.html>, fecha de consulta: 30 de mayo 2018.
- [8] <https://www.uni-muenster.de/Geoinformatics/>, fecha de consulta: 30 de mayo 2018.
- [9] <https://www.ulb.uni-muenster.de/>, fecha de consulta: 30 de mayo 2018.
- [10] <https://lodum.de/life/>, fecha de consulta: 30 de mayo 2018.
- [11] <https://www.upo.es/portal/impe/web/portada>, fecha de consulta: 12 de abril 2018.
- [12] <http://opendatasevilla.org/>, fecha de consulta: 12 de abril 2018.
- [13] <https://www.upo.es/estrategia/plan-estrategico/>, fecha de consulta: 12 de abril 2018.
- [14] <https://openspending.org/>, fecha de consulta: 12 de abril 2018.
- [15] <https://public.tableau.com/en-us/s/>, fecha de consulta: 12 de abril 2018.
- [16] <https://www.fundacionctic.org/proyectos/open-data-espana>, fecha de consulta: 12 de abril 2018.
- [17] <https://www.ua.es/>, fecha de consulta: 13 de abril 2018.
- [18] <https://web.ua.es/es/peua/plan-estrategico-ua-40.html>, fecha de consulta: 13 de abril 2018.
- [19] <https://github.com/UAdatos/opendata4u>, fecha de consulta: 13 de abril 2018.
- [20] <https://utc.ua.es/es/datos/>, fecha de consulta: 13 de abril 2018.
- [21] <https://www.sigua.ua.es/>, fecha de consulta: 13 de abril 2018.
- [22] <https://uaedf.ua.es/>, fecha de consulta: 13 de abril 2018.
- [23] <https://rua.ua.es/dspace/>, fecha de consulta: 13 de abril 2018.
- [24] <http://www.ba.unibo.it/>, fecha de consulta: 17 de abril 2018.
- [25] <http://www.exeter.ac.uk/research/openresearch/opendata/>, fecha de consulta: 17 de abril 2018.
- [26] <https://library.curtin.edu.au/help/faq/publishing-open-data/>, fecha de consulta: 17 de abril 2018.
- [27] <https://github.com/>, fecha de consulta: 17 de abril 2018.
- [28] <https://uwaterloo.ca/>, fecha de consulta: 17 de abril 2018.
- [29] <https://www.unam.mx/>, fecha de consulta: 14 de abril 2018.
- [30] <https://datosabiertos.unam.mx/>, fecha de consulta: 14 de abril 2018.
- [31] <http://www5.usp.br/>, fecha de consulta: 17 de abril 2018.
- [32] <http://www.unb.br/>, fecha de consulta: 17 de abril 2018.
- [33] <http://www.uda.cl/>, fecha de consulta: 17 de abril 2018.
- [34] <http://www.uchile.cl/>, fecha de consulta: 17 de abril 2018.
- [35] <http://uce.edu.ec/>, fecha de consulta: 17 de abril 2018.
- [36] <http://www.uncuyo.edu.ar/>, fecha de consulta: 17 de abril 2018.
- [37] <https://www.unam.edu.ar/>, fecha de consulta: 17 de abril 2018.
- [38] <http://www.uncuyo.edu.ar/transparencia/>, fecha de consulta: 17 de abril 2018.
- [39] <https://www.unam.edu.ar/index.php/transparencia>, fecha de consulta: 17 de abril 2018.
- [40] <http://estadisticasuniversitarias.me.gov.ar/#/home>, fecha de consulta: 7 de mayo 2018.
- [41] <https://collegescorecard.ed.gov/>, fecha de consulta: 7 de mayo 2018.
- [42] <https://www.ed.gov/>, fecha de consulta: 7 de mayo 2018.
- [43] <https://collegescorecard.ed.gov/data/>, fecha de consulta: 7 de mayo 2018.
- [44] <https://collegescorecard.ed.gov/data/documentation/>, fecha de consulta: 7 de mayo 2018.
- [45] <http://app.uni-muenster.de/>, fecha de consulta: 7 de mayo 2018.
- [46] <https://datos.ua.es/es/ficha-apps.html?idApp=1>, fecha de consulta: 9 de mayo 2018.
- [47] <https://data.ox.ac.uk/explore/science-area/>, fecha de consulta: 9 de mayo 2018.
- [48] <http://tools.southampton.ac.uk/room-finder/>, fecha de consulta: 9 de mayo 2018.
- [49] <https://chrome.google.com/webstore/detail/watmine/oehnjmhgpdjbiipeahdfmmbnldakogp>, fecha de consulta: 9 de mayo 2018.

Desarrollo

Introducción

Los avances tecnológicos de los últimos años y la competencia constante entre distintas empresas y comunidades permite contar hoy en día con una amplia variedad de herramientas útiles para llevar adelante proyectos de software.

Por tal motivo se debe llevar a cabo un análisis detallado de las herramientas para determinar cuáles se adaptan mejor a las necesidades del proyecto, la confección de modelos arquitectónicos adecuados, la selección de los lenguajes de programación y de los motores de bases de datos más convenientes para el proyecto en cuestión.

En este capítulo no solo se abordarán las tecnologías utilizadas para el prototipo desarrollado para esta tesina, sino también las diferentes comparativas entre la plataforma desarrollada y otras herramientas existentes para la disponibilización de datos abiertos como CKAN y Junar. Por último, se detallarán los beneficios que dispone el prototipo en cuestión con la finalidad de mostrar el por qué de un desarrollo alternativo.

Herramientas utilizadas para el prototipo

Ante el inminente crecimiento de las tecnologías informáticas abordamos una amplia investigación entre los diferentes lenguajes de programación, arquitecturas de aplicación y metodologías ágiles para determinar la eficacia y eficiencia en nuestro desarrollo.

A continuación se detallan las tecnologías utilizadas para el proyecto en cuestión junto con las alternativas con que se cuentan.

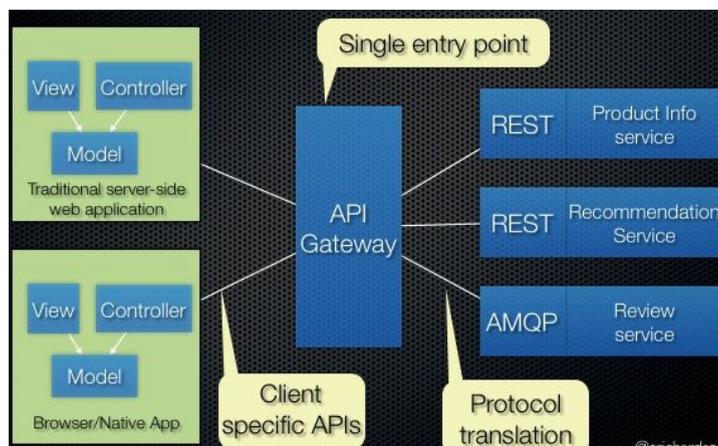
- **API REST**^{[1][2][3][4][5][6]}: Para definir que es una API REST, hay que separar los términos en dos: primero podemos definir API mediante sus siglas como "Application Programming Interface", o en español "Interfaz de programación de aplicaciones", es decir, un conjunto de funciones y procedimientos que cumplen una o muchas tareas con el fin de ser utilizado por otros sistemas. Permite que dos aplicaciones se comuniquen entre ellas mediante una capa de abstracción sin necesidad de duplicar código; es el mensajero que envía el requerimiento al servidor solicitado y luego devuelve la respuesta. Es importante no confundir servidor remoto con API, ya que este último es solo una parte del servidor que se encarga de recibir peticiones y enviar respuestas.

Existen APIs propias de un lenguaje, como Java APIs; o WEB APIs, como SOAP (Simple Object Access Protocol), RPC (Remote Procedure Call) o el más popular y últimamente más usado REST (Representational State Transfer).

El segundo término a desarrollar para entender qué es una API REST, es precisamente REST. Este concepto nacido en el año 2000, mediante la tesis doctoral de Roy Fielding^[7], describe una arquitectura de software definida especialmente para el desarrollo de servicios web. El mismo define la comunicación entre sistemas mediante el uso de las restricciones y

propiedades del protocolo HTTP, por ejemplo, utilizando los códigos estándar de HTTP se puede asumir que una respuesta 404 del servidor indicará que el recurso no existe o no se encuentra disponible; una respuesta 200 indicará que todo salió como se esperaba; y una respuesta 401 indicará que se necesitan más permisos para acceder al recurso. La implementación de un servicio REST sigue cuatro principios de diseño fundamentales:

- *Protocolo cliente/servidor sin estado*: cada petición HTTP contiene toda la información necesaria para ejecutarla, lo que permite que ni el cliente ni el servidor necesiten recordar el estado previo para satisfacerla. Sin embargo, en la práctica, muchas aplicaciones basadas en HTTP utilizan cookies y otros mecanismos para mantener el estado de la sesión.
 - *Un conjunto de operaciones bien definidas que se aplican a todos los recursos de información*: HTTP en sí define un conjunto pequeño de operaciones, las más importantes son POST, GET, PUT y DELETE. Con frecuencia estas operaciones se equiparan a las operaciones CRUD en bases de datos.
 - *Los objetos en REST siempre se manipulan a partir de la URI*: Es la URI el identificador único de cada recurso de ese sistema REST. La URI facilita el acceso a la información para su modificación o eliminación. También para compartir su ubicación exacta con terceros.
 - *El uso de hipermédios, tanto para la información de la aplicación como para las transiciones de estado de la aplicación*: la representación de este estado en un sistema REST son típicamente HTML, XML o JSON. Como resultado de esto, es posible navegar de un recurso REST a muchos otros, simplemente siguiendo enlaces sin requerir el uso de registros u otra infraestructura adicional.
- **Middleware API Gateway**^{[8][9][10]}: La API Gateway es responsable del enrutamiento de solicitudes, la composición y la traducción de protocolos. Además maneja estas solicitudes de dos maneras: algunas son simplemente proxeadas/enrutadas al servicio apropiado, mientras que otras se gestionan dividiéndose en múltiples servicios. La API Gateway se encarga también de la implementación de seguridad, por ejemplo la autorización de un usuario sobre un recurso; y de enmascarar fallas en los servicios de back-end devolviendo datos en caché o por defecto.

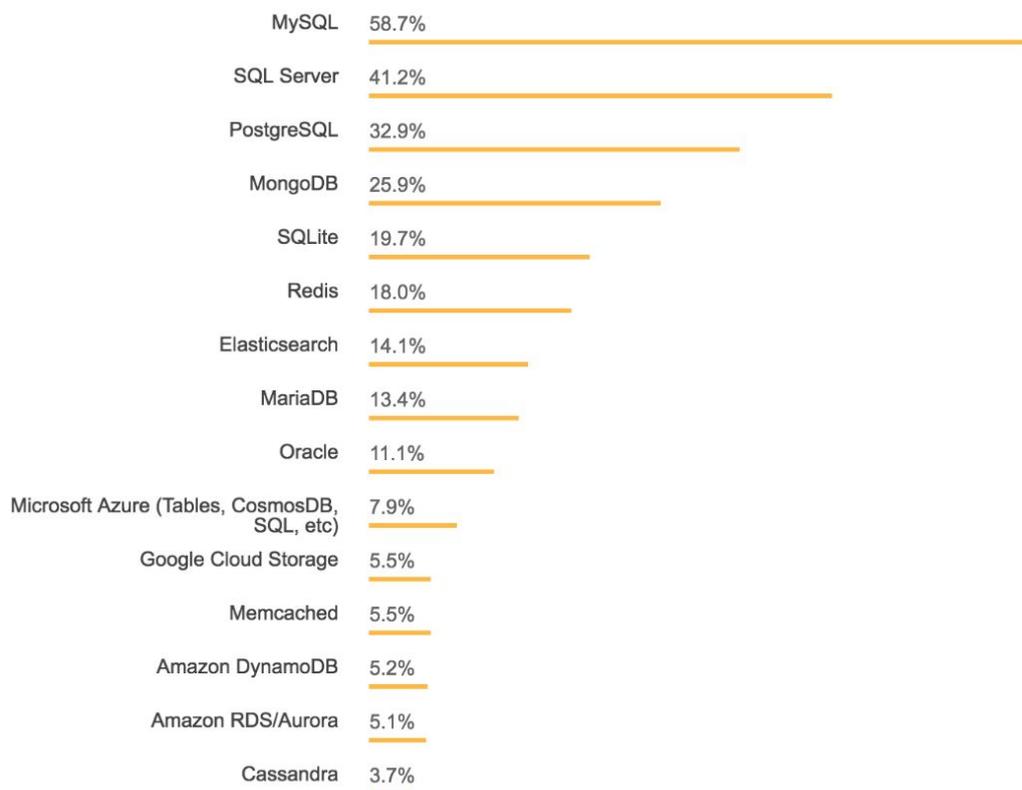


Arquitectura con un API Gateway.

- **Base de datos relacional^[11]:** Un base de datos relacional es una recopilación de elementos de datos con relaciones predefinidas entre ellos. Estos elementos se organizan como un conjunto de tablas con columnas y filas. Las tablas se utilizan para guardar información sobre los objetos que se van a representar en la base de datos. Cada columna de una tabla guarda un determinado tipo de dato y un campo almacena el valor real de un atributo. Las filas de la tabla representan una recopilación de valores relacionados de un objeto o entidad. Cada fila de una tabla podría marcarse con un identificador único denominado clave principal, mientras que las filas de varias tablas pueden relacionarse con claves extranjeras. El acceso a estos datos se puede obtener de diferentes formas sin reorganizar las tablas de la base de datos. Los aspectos más importantes son los siguientes:
 - *SQL* es el lenguaje principal. Se convirtió en un estándar del American National Standards Institute (ANSI) en 1986. Todos los motores de bases de datos relacionales populares soportan *SQL* de ANSI estándar. *SQL* se utiliza para agregar, actualizar o eliminar filas, recuperar subconjuntos de datos para aplicaciones de análisis y procesamiento de transacciones y para administrar todos los aspectos de la base de datos.
 - La *integridad* de los datos es la totalidad, precisión y coherencia general de los datos. Las bases de datos relacionales utilizan un conjunto de restricciones para aplicar la integridad de los datos en la misma. Esto incluye claves principales, claves externas, restricción "Not NULL", restricción "Unique", restricción "Default" y restricciones "Check". Estas restricciones de integridad ayudan a aplicar reglas de negocio en datos de las tablas para garantizar la precisión y fiabilidad de los mismos. Además de las anteriores, la mayoría de las bases de datos relacionales también permiten la integración de código personalizado en disparadores que se ejecutan en función de una acción en la base de datos.
 - Una *transacción* de base de datos es una o más sentencias *SQL* que se ejecutan como una secuencia de operaciones que forman una unidad lógica única de trabajo. Las transacciones proporcionan una proposición "todo o nada", por lo que la transacción completa se debe completar como una sola unidad y se debe escribir en la base de datos; de lo contrario, ninguno de los componentes individuales de la transacción debería pasar. En la terminología de bases de datos relacionales, una transacción genera un COMMIT o un ROLLBACK. Cada transacción se trata de forma coherente y fiable independiente de otras transacciones.
 - Todas las transacciones de base de datos deben ser conformes a *ACID* (atómicas, coherentes, aisladas y duraderas) para garantizar la integridad de los datos. La *atomicidad* requiere que la transacción completa se ejecute correctamente; si una parte de la transacción falla, toda ella queda invalidada. La *coherencia* exige que los datos escritos en la base de datos como parte de la transacción cumplan todas las reglas definidas, así como las restricciones, cascadas y disparadores. El *aislamiento* es crítico para lograr el control de concurrencia y se asegura de que cada transacción sea independiente por sí misma. La *durabilidad* requiere que todos los cambios realizados en la base de datos sean permanentes una vez que la transacción se complete correctamente.

Los motores de bases de datos más utilizados en el mercado son: PostgreSQL, MySQL^[12], MariaDB^[13], SQL Server^[14] y Oracle^[15].

- PostgreSQL^[16]**: Nacido en 1986 en la Universidad de California en Berkeley, es un potente sistema de base de datos relacional de código abierto que utiliza y amplía el lenguaje SQL combinado con muchas características que almacenan y escalan de forma segura las cargas de trabajo de datos más complicadas. PostgreSQL se ha ganado una sólida reputación por su arquitectura comprobada, confiabilidad, integridad de datos, sólido conjunto de características, extensibilidad y la dedicación de la comunidad de código abierto detrás del software para entregar constantemente soluciones eficaces e innovadoras. PostgreSQL se ejecuta en todos los principales sistemas operativos, cumple con las propiedades ACID desde 2001 y tiene complementos potentes, como el popular extensor de base de datos geoespaciales PostGIS^[17], el cual agrega soporte para objetos geográficos, permitiendo hacer consultas sobre ubicaciones. Se ha demostrado que PostgreSQL es altamente escalable tanto por la gran cantidad de datos que puede gestionar como por la cantidad de usuarios simultáneos que puede acomodar. Hay clústeres PostgreSQL activos en entornos de producción que administran muchos terabytes de datos y sistemas especializados que administran petabytes. Según la encuesta realizada por Stack Overflow^[18], PostgreSQL es la tercera base de datos más utilizada y realizando la comparativa con encuestas de años anteriores está en crecimiento.



Encuesta realizada por Stack Overflow 2018.

- BBDD No relacional (NoSQL)^{[19][20]}**: Son bases de datos no relacionales optimizadas para modelos de datos sin esquema y de desempeño escalable. Estos tipos de bases de datos están optimizados para aplicaciones que requieren grandes volúmenes de datos, baja latencia y modelos de datos flexibles, lo que se logra mediante la flexibilización de algunas de las restricciones de coherencia de datos en las bases de datos relacionales tradicionales. Utilizan una variedad de modelos de datos como los almacenes de valor clave en memoria,

de gráficos, de documentos y en columnas. Comparadas con las bases de datos relacionales, las NoSQL son más escalables y tienen una performance superior, en especial en grandes volúmenes de datos estructurados, semi estructurados y desestructurados que cambian rápidamente, también poseen programación orientada a objetos que es fácil de usar y flexible como lenguaje de base de datos. En cuanto a comparación de consultas, en los RDBMS, los datos se pueden consultar de manera flexible, pero las consultas son relativamente costosas y no se escalan bien cuando hay mucho tráfico; por otro lado en una base de datos NoSQL, los datos se pueden consultar de manera eficiente en un número limitado de maneras, fuera de los cuales es más costoso y lento hacer consultas.

- **MongoDB^[21]**: MongoDB fue fundada en 2007 por Dwight Merriman, Eliot Horowitz y Kevin Ryan, el equipo detrás de DoubleClick. En la empresa de publicidad en Internet DoubleClick (ahora propiedad de Google), el equipo desarrolló y usó muchos almacenes de datos personalizados para solucionar las deficiencias de las bases de datos existentes. El negocio atendió 400,000 anuncios por segundo, pero a menudo tuvo problemas de escalabilidad y agilidad. Frustrado, el equipo se inspiró para crear una base de datos que abordara los desafíos que enfrentaba en DoubleClick. Hoy en día, MongoDB es la plataforma de base de datos moderna líder de propósito general, diseñada para liberar el poder del software y los datos para los desarrolladores y las aplicaciones que construyen.
- **JavaScript^{[22][23][24]}**: Es un lenguaje de programación que se utiliza principalmente en navegadores web para desarrollar páginas dinámicamente, pero también se utiliza del lado del servidor. Este lenguaje permite a los desarrolladores manipular el contenido de las páginas a través del DOM; manejar los datos con AJAX e IndexedDB; crear gráficos con Canvas; interactuar con el dispositivo que ejecuta el navegador utilizando varias APIs, etc. El creciente uso de APIs en los navegadores lo convierte en uno de los lenguajes de programación más usados en el mundo. JavaScript ha tenido éxito de la mano de Node.js para los desarrollos del lado del servidor. Esta plataforma provee un entorno completo de ejecución JavaScript fuera del navegador, capaz de ser usado en cualquier sistema operativo (Linux, MacOS y Windows). Si bien no es el único, es el más usado últimamente para correr JavaScript fuera de los navegadores. Esto permite que JavaScript sea utilizado como un lenguaje de scripting para automatizar ciertos procesos en la PC, así como construir HTTP completamente funcional y servidores Web Sockets.

JavaScript se diseñó con una sintaxis similar a C y adopta nombres y convenciones del lenguaje de programación Java. Sin embargo, Java y JavaScript tienen semánticas y propósitos diferentes.

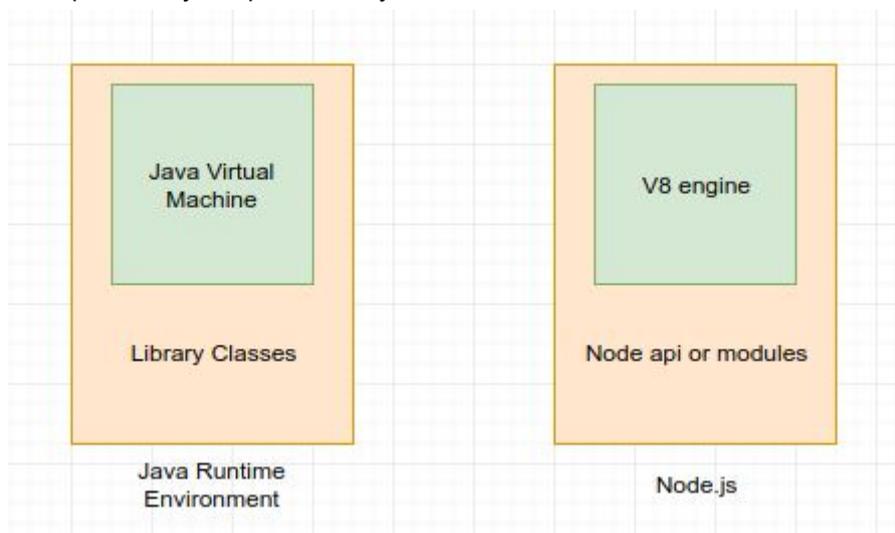
El surgimiento de aplicaciones web más sofisticadas acompañadas de una velocidad para navegar bastante lenta, originó la necesidad de concebir un lenguaje de programación que se ejecutara en el mismo navegador, con el objetivo de optimizar el tiempo de navegación de los usuarios; por ejemplo, si un usuario ingresara de forma incorrecta su información en un formulario, no tendría que esperar mucho tiempo para detectar y corregir los errores. Por este motivo, el programador Brendan Eich, quien trabajaba en la empresa estadounidense de software Netscape, comenzó a diseñar la idea de JavaScript. La versión inicial tuvo mucho éxito y fue bien recibida, por lo que Netscape lanzó su Netscape Navigator 3.0 con una versión JavaScript 1.1.

Desde su aparición, JavaScript siempre fue utilizado de forma masiva por la mayoría de los sitios web. La aparición de Flash disminuyó su popularidad, ya que este permitía realizar

algunas acciones, como mostrar gráficos y vectores en movimiento, imposibles de llevar a cabo mediante JavaScript. Sin embargo, la aparición de las aplicaciones AJAX, programadas con JavaScript, le ha devuelto la popularidad. Como resultado de esto hubo una proliferación de un conjunto de frameworks y librerías de ámbito general, mejorando las prácticas de programación con JavaScript y aumentando el uso del lenguaje.

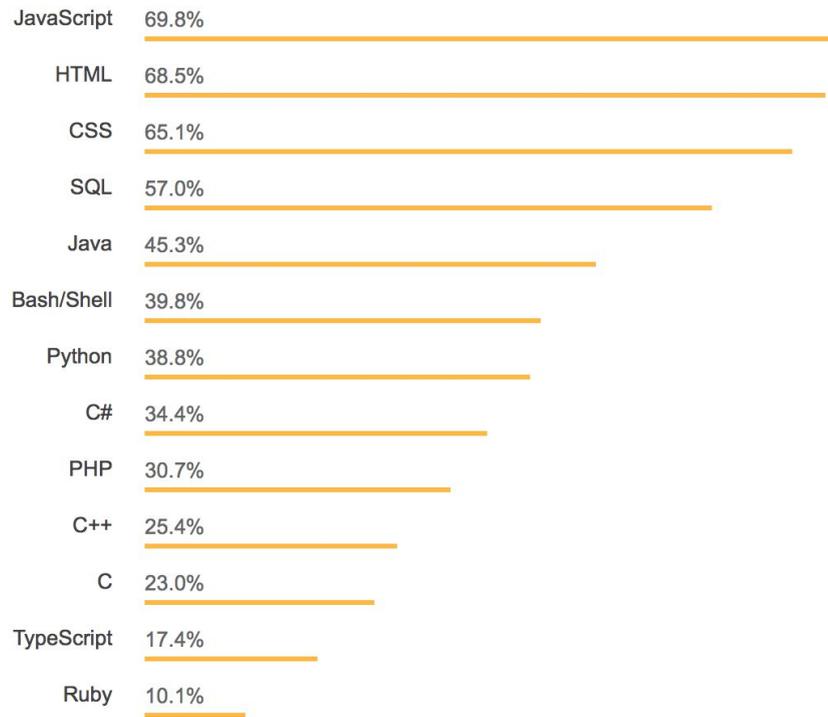
En cuanto a las limitaciones, JavaScript fue diseñado para ejecutarse en un entorno muy limitado que permite a los usuarios confiar en la ejecución de los scripts. De esta forma, no pueden comunicarse con recursos que no pertenezcan al mismo dominio desde el que se descargó el script. Tampoco pueden cerrar ventanas que no hayan abierto esos mismos scripts. Las ventanas no pueden ser demasiado pequeñas, ni demasiado grandes, ni colocarse fuera de la vista del usuario (aunque los detalles concretos dependen de cada navegador). Además, no pueden acceder a los archivos de la computadora del usuario (ni en modo lectura, ni en modo escritura) y tampoco pueden leer o modificar las preferencias del navegador. Por último, si su ejecución dura demasiado tiempo, por ejemplo, por un error de programación, el navegador informa al usuario de que un script está consumiendo demasiados recursos y le da la posibilidad de detener su ejecución.

- **Node.js**^{[25][26]}: Es un entorno de ejecución para JavaScript construido con el motor de JavaScript V8 de Chrome. Node.js utiliza un modelo de operaciones E/S sin bloqueo y orientado a eventos, que lo hace liviano y eficiente; además está orientado a evento asíncronos. Esta última característica contrasta con el modelo de concurrencia más común hoy en día, donde se usan hilos del Sistema Operativo. Las operaciones de redes basadas en hilos son relativamente ineficientes y muy difíciles de usar. Además, los usuarios de Node están libres de preocupaciones sobre el bloqueo del proceso, ya que este no existe. Casi ninguna función en Node realiza I/O directamente, por lo que el proceso nunca se bloquea. Debido a que no hay bloqueo es muy razonable desarrollar sistemas escalables en Node.



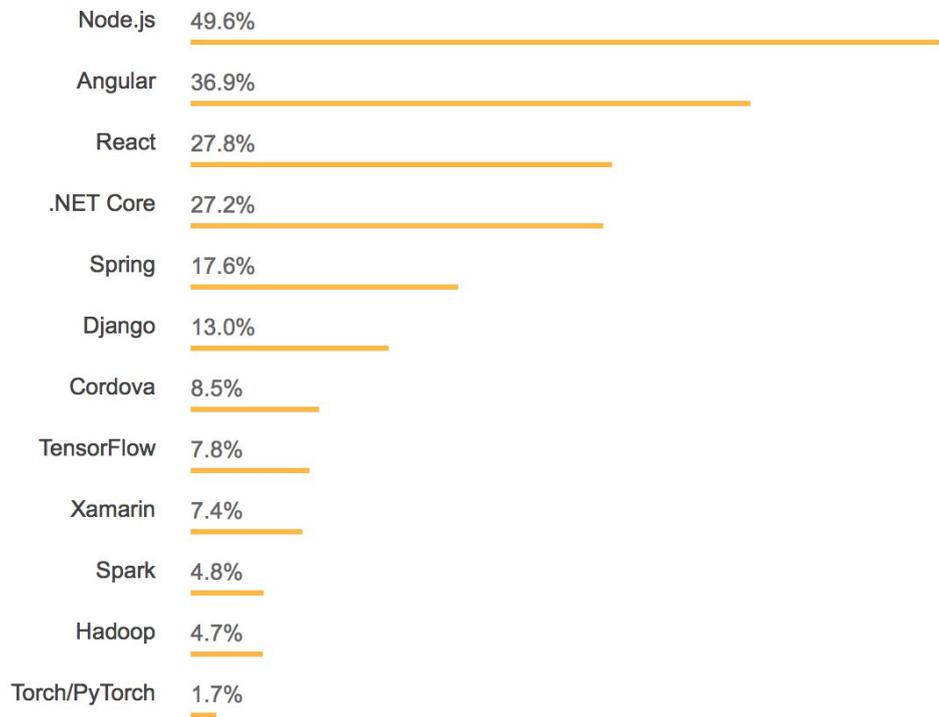
Comparativa de un servidor utilizando Java vs Node.js

Según encuestas de Stack Overflow, la comunidad más grande de desarrolladores, Javascript es, por sexto año consecutivo, el lenguaje más utilizado^[27].



Encuesta realizada por Stack Overflow 2018.

Otra encuesta realizada por la misma comunidad^[27], muestra a Node.js liderando el uso en la categoría "Frameworks, Librerías y Herramientas".



Encuesta realizada por Stack Overflow 2018.

- **npm^[28]**: Es un administrador de paquetes para Node.js. Fue lanzado en el año 2010 y hoy en día se convirtió en el registro de software más grande del mundo, con aproximadamente 3 mil millones de descargas por semana. Npm contiene más de 600.000 paquetes y se utiliza mayormente para contribuciones open source, pero también tiene la opción de administrar paquetes de desarrollo privados para organizaciones. Se instala por defecto con Node.js, y tiene una CLI (command line interface) muy simple, desde la cual se puede interactuar con npm, descargando, publicando y actualizando paquetes.

- **Sails.js^{[29][30]}**: Fue desarrollado por Mike McNeil para ayudar a su equipo a crear proyectos escalables de Node.js para clientes empresariales y de startups. Desde su lanzamiento en 2012, Sails se ha convertido en uno de los marcos de aplicaciones web más utilizados en el mundo. Facilita la creación de aplicaciones Node.js personalizadas y de nivel empresarial. Está diseñado para parecerse a la arquitectura MVC de frameworks como Ruby on Rails, pero con soporte para el estilo de aplicaciones web y API más moderno y orientado a los datos.

- **AngularJs^{[31][32][33]}**: Es un framework de JavaScript de código abierto, mantenido por Google^[34], que se utiliza para crear y mantener aplicaciones web de una sola página. Dicho framework respeta algunas características como:
 - **Mobile First**: Es una filosofía que propone analizar el funcionamiento de una aplicación en un dispositivo móvil antes de su versión web o desktop. AngularJs aporta a esto ya que posee un sistema de routing específico para mantener el código de la aplicación liviano con lo cual disminuye el tiempo de espera por parte de los usuarios. Además incluye características para soportar gestos, toques y giros en el celular, siempre con un óptimo desempeño debido a que utiliza poca memoria.
 - **ECMAScript V6^[35]**: AngularJs está escrito sobre el último estándar de JavaScript, lo cual conlleva una mejora significativa en la ejecución de sus módulos y nuevas características para elevar las prestaciones del producto.
 - **Desarrollo Flexible**: A pesar de que AngularJs tiene soporte para ECMAScript V6 y TypeScript, también ofrece retrocompatibilidad y funcionalidad con ECMAScript V5, Dart^[36], CoffeeScript^[37] u otros lenguajes que compilen JavaScript.
 - **Velocidad y Desempeño**: AngularJs ofrece la detección de cambios de estado a un velocidad bastante más elevada a la media ya que utiliza el estándar más moderno de JavaScript.
 - **Simple y Expresivo**: Permite escribir código limpio y natural, reduciendo la complejidad para el equipo de desarrollo, haciendo el código altamente legible y entendible.
 - **Routing Comprensivo**: Transiciones entre URLs, rutas hermana y rutas anidadas.
 - **Animaciones**: Por defecto AngularJs posee una API de animación que permite crear secuencias complejas de animación.

- **Inyección de dependencias jerárquica:** AngularJs posee la característica de disponer el código de forma modular, mantenible y extensible a través del tiempo mediante la inyección de dependencias, también permite probar el código a través de esta característica.
- **Soporte para componentes web:** Ofrece la creación e integración de componentes web como Polymer^[38] y X-Tag^[39], entre otros.
- **Internacionalización y Accesibilidad:** Permite usar características de traducción, pluralización y reglas de género.

Con el surgimiento de las necesidades especiales por parte de los desarrolladores de software dedicados al lado del cliente y el creciente avance de JavaScript, surgieron frameworks que implementan el patrón MVC (Modelo, Vista, Controlador) con el fin de separar conceptos y mantener cierto orden en los productos. El más conocido es BackboneJs^[40], que fue creado en 2010 por Jeremy Ashkenas, también creador de CoffeeScript, y depende de otras dos librerías: jQuery^[41] y Underscore.js^[42]. Es usado por múltiples startups como Pinterest, Foursquare, AirBnB y Trello. BackboneJs permite crear una aplicación rápidamente aunque en ocasiones es complicado de utilizar. La mayoría de los desarrolladores eligen este framework ya que parece la opción más segura, está vigente hace más tiempo, cuenta con mucha documentación y tiene mantenimiento y respaldo por parte de una gran comunidad.

El surgimiento de AngularJs tuvo como primer medida el desplazamiento de varios de los frameworks que eran furor hasta el momento del lanzamiento de BackboneJs. Si bien su primera versión es de 2009, se hizo muy popular a finales de 2012 y fue recién en 2013 cuando estuvo en pleno auge. Comenzó a hablarse de la nueva technology stack MEAN (MongoDB/Mongoose + ExpressJS + AngularJS + Node.js), como en su época había sido el turno de LAMP (Linux + Apache + MySQL + PHP). Como punto a su favor se encuentra la simplicidad al momento de crear tests unitarios y end-to-end con Jasmine^[43] y Karma^[44], algo que suele ser un poco difícil al principio. Además, permite extender el vocabulario HTML con directivas y atributos, manteniendo la semántica y sin necesidad de emplear librerías externas como jQuery o Underscore.js.

- **GulpJs^{[45][46][47][48]}:** Con el auge de los nuevos frameworks JavaScript como AngularJs, BackboneJs, EmberJs^[49], ReactJs^[50] y Polymer, entre otros, surge la necesidad de contar con herramientas para gestionar proyectos como GulpJs y GruntJs. De manera análoga en Java^[51], existe Maven^[52]; o en Groovy^[53], Gradle^[54].

GulpJs es una herramienta para la gestión y automatización de tareas en JavaScript. Estas tareas son de uso común para los desarrolladores. El único requisito para utilizar GulpJs es tener instalado Node.js. Algunas de las tareas más frecuentes de son:

- Compilación de CSS y Javascript preprocesado.
- Concatenación.
- Minificación.
- Lanzamiento de un servidor para la recarga automática en el navegador.
- Creación de una build para despliegue.
- Ejecución de tests unitarios.

- Ejecución de herramientas para detectar errores en el código.
- Gestión del código en un repositorio.

A diferencia de otras herramientas de este tipo, como GruntJs^[55], GulpJs utiliza streams para la ejecución de tareas. Un stream es un flujo de datos para el intercambio de información. GulpJs utiliza el módulo Stream^[56] de Node.js. Para realizar operaciones sobre un stream se utilizan pipes. Estos son un mecanismo para la comunicación y sincronización entre procesos, y cuentan con diferentes características como por ejemplo, que se encuentran basadas en el patrón productor/consumidor, están implementadas de forma muy eficiente en los sistemas operativos, inician todos los procesos al mismo tiempo y atienden automáticamente los requerimientos de lectura de datos para cada proceso cuando los datos son escritos por el proceso anterior.

GulpJs lanzada en 2014, no fue la primera herramienta relevante de este tipo: GruntJs surgió dos años antes, en 2012. Aunque GruntJs sigue gozando de una fuerte popularidad, gran cantidad de plugins desarrollados y amplia comunidad de desarrolladores; GulpJs es superior en varios aspectos, lo cual hace que ya en este momento, y en el futuro, GulpJs sea la herramienta de automatización de tareas predominante.

Sus diferencias más importantes son: en primer lugar, GruntJs tiene un enfoque de configuración sobre código, mientras que GulpJs tiene un enfoque de código sobre configuración; esto hace que GulpJs sea mucho más conciso y claro, mientras que GruntJs es más verboso. Y en segundo lugar, GulpJs es mucho más eficiente por su concepción basada en el uso de streams que implica operaciones de lectura y escritura en memoria, mientras que en GruntJs, se realizan varias lecturas y escrituras en disco para encadenar varias operaciones en una tarea.



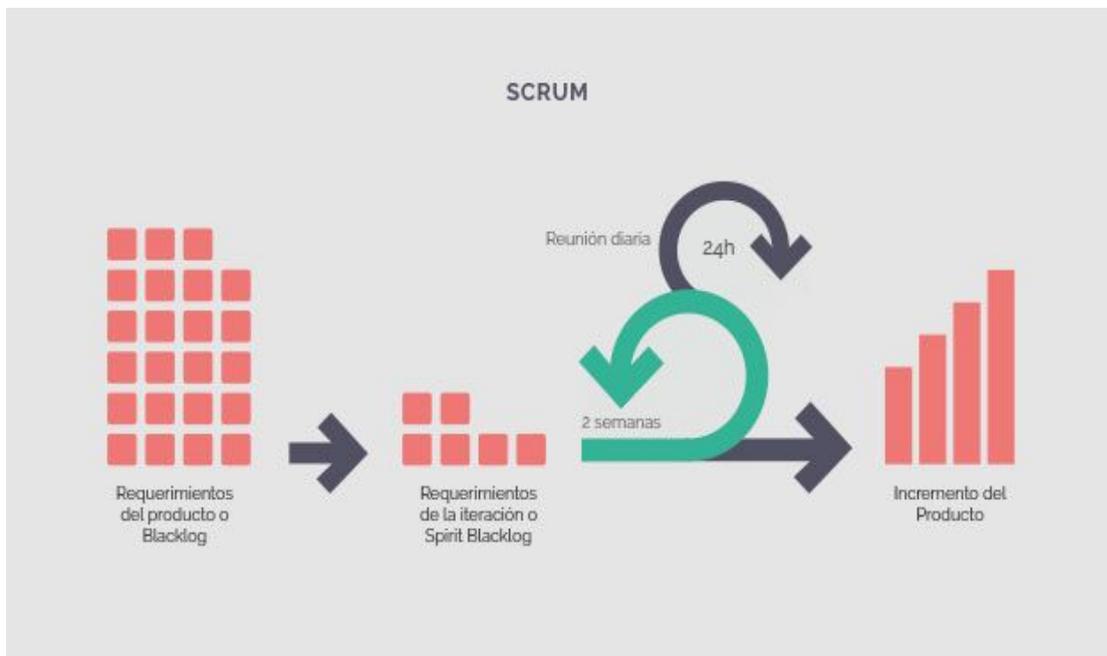
Demostración del flujo de GulpJs para la lectura-escritura de los archivos a automatizar.

- **Metodologías ágiles^{[57][58]}:** Lo primero a definir es el concepto de metodologías ágiles. Por metodologías ágiles entendemos a aquellas metodologías de gestión que permiten adaptar la forma de trabajo al contexto y naturaleza de un proyecto, basándose en la flexibilidad y la inmediatez, y teniendo en cuenta las exigencias del mercado y los clientes. Los pilares fundamentales de las metodologías ágiles son el trabajo colaborativo y en equipo.

Los principales beneficios de utilizar estas metodologías son:

- Reducción de costos.
- Rapidez en la entrega de proyectos.
- Trabajo en equipo y compromiso de todos los miembros del equipo de trabajo.
- Mayor calidad en el trabajo y en el producto o servicio final.

Scrum es una de estas metodologías ágiles y consiste en un proceso en el que se aplican de forma regular un conjunto de buenas prácticas para trabajar colaborativamente y obtener el mejor resultado posible de un proyecto. Estas prácticas se apoyan unas a otras y su selección tiene origen en un estudio de la manera de trabajar de equipos altamente productivos.



Ciclo de ejecución de la metodología Scrum.

Esta metodología consiste en realizar períodos de entrega de trabajo, cuya base principal es su ciclo de vida iterativo e incremental. El ciclo de vida iterativo o incremental es aquel en que se va liberando el producto por partes, periódicamente, iterativamente, poco a poco y además, cada entrega es el incremento de funcionalidad respecto a la anterior. Cada periodo de entrega se conoce con el nombre de Sprint.

El segundo pilar importante del Scrum son las revisiones. Su valor reside en que las reuniones son la base para lograr transparencia y comunicación, y posibilitan la retroalimentación del equipo. Se pautan reuniones de revisión:

- **Reunión de planificación del sprint:** Al principio de cada sprint se decide lo que se va a realizar en el mismo.
- **Reunión diaria:** Máximo de 15 minutos para conceptualizar lo realizado el día anterior, lo que se realizará ese día y qué problemas se han encontrado.
- **Reunión de revisiones del sprint:** Al final de cada sprint se plantean las tareas realizadas y las que no.

- **Retrospectiva del sprint:** Esto se lleva adelante al final de cada sprint y sirve para que los implicados en el proyecto expresen sus impresiones del sprint. Se utiliza para mejorar el proceso abordado.

Plataformas de publicación de datos abiertos

Comúnmente a las herramientas que se utilizan en las iniciativas de datos abiertos para gestionar su apertura, se las conoce como “Catálogo de datos”. Es importante aclarar que cumplen con funcionalidades más amplias, como proporcionar el soporte necesario de los módulos que integran el catálogo y que finalmente abastecen una catalogación de los datos, por lo que sería conveniente mencionarlas como “Plataformas de publicación de datos”^[59].

Las plataformas de publicación de datos deben contar con ciertas características para lograr un mayor grado de optimización en su funcionamiento. Dichas características son:

- **Catálogo de datos:** Es el elemento principal para cualquier plataforma de publicación de datos, ya que es el punto de contacto entre la entidad que disponibiliza la información y el consumidor final. El catálogo es el responsable de organizar el conjunto de datos y facilitar la reusabilidad de la información dentro de los parámetros mencionados de dato abierto. Este elemento debe contar con algunas características como:
 - **Soporte para metadatos:** Los metadatos proporcionan información adicional sobre los conjuntos de datos disponibles. Esa información resulta de gran valor a la hora de organizar, clasificar, categorizar y buscar a través de grandes volúmenes de información, lo que permite agilizar y establecer un análisis sobre los datos de interés.
 - **Compatibilidad con estándares de datos abiertos:** El cumplimiento de los estándares proporcionará ventajas en términos de interoperabilidad, facilitando por ejemplo, la federación y agregación de conjuntos de datos entre distintos catálogos de datos relacionados.
 - **Soporte avanzado para Linked Data Open:** Para apuntar a la escalabilidad de la plataforma es importante que cuente con soporte para la gestión y publicación de datos utilizando tecnologías Linked Data^[60] y Web Semántica^[61]. Para lograr este tipo de desarrollo se deben respetar requisitos específicos^[62] de datos enlazados, por lo que resulta conveniente optar por una integración con herramientas dedicadas a esta funcionalidad. Sin embargo, algunos de los aspectos a lograr son: soporte a representaciones semánticas como RDF/XML y JSON-LD y la posibilidad de realizar consultas dinámicas a través de SPARQL^[63].
- **Servicios de datos:** La plataforma debe facilitar la interacción con el consumidor final de los datos disponibilizados. Para esto es importante tener en cuenta los siguientes aspectos:
 - **Acceso a los datos:** Principalmente se debe contar con un enlace directo a los datos disponibilizados, ya sea a través de enlaces externos como descargas directas dentro del catálogo propio. Es recomendable abastecer APIs, idealmente de tipo REST, de acceso abierto para agilizar la interacción programática y utilizar formatos de tipo JSON y XML.

- **Búsqueda y clasificación:** La funcionalidad principal de cualquier plataforma interactiva es la búsqueda sobre el contenido de la misma. Es importante establecer un sistema de búsqueda acorde a las necesidades del consumidor de la información con el fin de agilizar su experiencia con los conjuntos de datos disponibles.
- **Actualización:** Este elemento corresponde a una característica propia de los datos. La continua evolución de la información establece una constante actualización de la misma, lo que genera la necesidad de proporcionar un sistema de versionado de la información y, por consiguiente, mantener un historial de las mismas.
- **Visualización:** Las visualizaciones de la información en el catálogo de datos permiten interacciones más amigables y comprensibles al consumidor del portal, haciéndolo más inclusivo y apuntado a diferentes tipos de públicos.
- **Galería de aplicaciones:** Es un elemento adicional que permite un doble beneficio: por un lado, aumenta la utilidad y el valor del dato reutilizado a partir del consumo de los nuevos servicios; por otro, incentiva a posibles reutilizadores de datos a generar proyectos innovadores.
- **Servicios informativos:** Esta característica apunta a la difusión de la iniciativa de datos abiertos y a captar el interés del público para que sea colaborativo y participativo. Algunos elementos corresponden a:
 - **Descripción de la iniciativa de apertura de información:** El objetivo de este elemento es introducir al usuario final del catálogo de datos en la iniciativa de datos abiertos o el concepto de apertura de información, que si bien ha ido incrementando su divulgación en los últimos años, en varios ámbitos sigue siendo nueva. Se deben explicar las motivaciones de la iniciativa, sus características básicas y el plan de acción de la apertura de datos.
 - **Noticias y novedades:** Es importante lograr una interacción constante con el usuario y mantenerlo informado sobre los actualizaciones que se producen en la plataforma o sobre temas relacionados con datos abiertos. Es conveniente adquirir un canal de comunicación que permita obtener un feedback por parte de quien utiliza la plataforma para lograr mejoras y avances en el proyecto. Además, es buena idea incluir otros canales por los cuales comunicar y establecer una conexión como por ejemplo, las redes sociales.
 - **Documentación y ayuda:** Es importante disponibilizar al consumidor de la información toda la documentación correspondiente a los estándares y buenas prácticas utilizadas en el desarrollo de la herramienta.
 - **Términos de uso y licencias:** Toda iniciativa de datos abiertos necesita definir un marco legal único, claro y transparente para proporcionar la sostenibilidad y fiabilidad requerida.
 - **Estadísticas de uso:** Este elemento aporta un mayor grado de transparencia sobre el uso y apertura de la plataforma, ya que contempla no solo el acceso y uso de la misma, sino también las descargas e interacciones con los datos y la gestión sobre ellos, sus actualizaciones, las frecuencias y los formatos utilizados, entre otros.

- **Servicios para la comunidad:** Esta característica amplía la de servicios informativos que se orienta a la comunicación constante con la comunidad de reutilizadores y consumidores de los datos. Esta herramienta debe contar con secciones como:
 - **Comentarios, ideas y sugerencias:** Es ideal disponibilizar un canal de comunicación sobre los conjuntos de datos o recursos dentro de la propia plataforma con el fin de que el usuario pueda realizar sugerencias o comentarios sobre los mismos.
 - **Peticiones de datos:** Uno de los puntos más importantes de las iniciativas de datos abiertos es poner a disposición del usuario los datos que necesitan para su beneficio (dentro de los marcos establecidos de datos abiertos), por lo que se debe contar con una sección o canal donde se puedan hacer pedidos sobre datos o información que no se encuentre disponible. Esto enriquece debido a la interacción entre la entidad disponibilizadora y el consumidor, y a la generación de nuevos recursos o conjuntos de datos.

A medida que se fue profundizando el concepto de iniciativa de datos abiertos, ha incrementado la cantidad de herramientas orientadas a la publicación de los mismos. Entre las más destacadas se encuentran:

- **CKAN:** Una herramienta promovida por la asociación Open Knowledge y cuenta con un flujo de trabajo similar a un gestor de contenidos tradicional (CMS) orientado específicamente a la carga y gestión de datos. CKAN cuenta con un amplio rango de características, entre ellas se destacan:
 - **Publicación y gestión de datos:** carga de datos a través una subida directa de archivos o utilizando servicios web y la gestión de categorías y permisos.
 - **Búsqueda y reutilización:** búsquedas a través de palabras clave o con filtros sobre metadatos.
 - **Metadatos:** amplio conjunto de metadatos disponibles por defecto, incluyendo metadatos geoespaciales y la posibilidad de incorporar nuevos metadatos personalizados.
 - **Herramientas de comunidad:** posibilidad de compartir conjuntos de datos a través de redes sociales, suscripción a contenidos y actualizaciones de conjuntos de datos y herramientas para obtener feedback de la comunidad.
 - **Visualización:** posibilidad de previsualizar los datos estructurados a través de tablas dinámicas o gráficos básicos. Geolocalización de datos que cuenten con algún componente geográfico. En el caso de los gráficos y mapas, se deben utilizar herramientas externas para su correcta visualización.

Además, CKAN permite la extensión de funcionalidades y modificaciones en la apariencia del catálogo de datos a partir del desarrollo de temas. La plataforma de CKAN está disponible de forma gratuita como software libre, pero sus creadores también ofrecen servicios adicionales

de desarrollo, alojamiento, soporte y formación a los que hay que contactar previamente para obtener un presupuesto.

- **Junar:** Es una plataforma de datos abiertos en la nube que facilita la publicación de datos por parte de gobiernos, empresas u otras organizaciones. A través de Junar se pueden seleccionar los conjuntos de datos que se desean publicar, así como el tiempo y la forma en la que se van a presentar al público. Entre los servicios destacados que ofrece se encuentran:
 - **Recolección de datos:** Los datos pueden ser introducidos a través de archivos individuales o enlaces en la web. Además ofrece soporte para los formatos más comunes y permite hacer un filtrado previo de la información que contiene.
 - **Visualizaciones de los datos:** Uno de los puntos fuertes de la plataforma son las visualizaciones y la flexible gestión de las mismas. Además, permite hacer un seguimiento estadístico del uso y visibilidad que tienen las visualizaciones gestionadas desde la herramienta.
 - **Publicación:** el proceso de publicación en Junar tiene en cuenta los procesos internos de la organización y permite asignar los permisos necesarios para cada usuario a la hora de recopilar, mejorar y publicar los datos. Este método de gestión de permisos incluye la gestión de datos privados para uso exclusivamente interno. Por otro lado, Junar ofrece también una Interfaz de Programación (API) integrada para que los desarrolladores de servicios y aplicaciones puedan reutilizar los datos más fácilmente.
 - **Análisis del uso:** permite hacer seguimiento en tiempo real del uso de los datos gracias a las estadísticas proporcionadas por Junar.

Esta herramienta es comercial y tiene un costo mensual para su desarrollo básico, lo que supone un costo adicional para la incorporación de nuevos módulos o extensiones.

Tanto CKAN como Junar son los casos de estudio específicos de este documento ya que son las herramientas más utilizadas del mercado en la República Argentina, pero el abanico de plataformas de publicación de datos es más amplia. Otras alternativas son:

- **DKAN:** es un clon de CKAN pero basado en la plataforma de Drupal^[64], lo que le confiere una gran flexibilidad a la hora de aprovechar la extensa comunidad que existe detrás de dicha herramienta y la tecnología de desarrollo que utiliza.
- **data.gov.uk To Go**^[65]: es un kit preparado por el Gobierno de Reino Unido para poner a disposición del público su plataforma de publicación de datos (que integra Drupal y CKAN) de forma que cualquiera pueda desplegar una plataforma similar preocupándose únicamente de adaptar la apariencia externa final.
- **DataPress**^[66]: es un proyecto del Open Data Institute basado en tecnología de código abierto para la gestión de los datos de la administración en la nube sobre la base de Wordpress y CKAN y realizada por el mismo equipo de data.gov.uk
- **OpenDataSoft**^[67]: es una plataforma basada en servicios en la nube que cuenta ya con un cierto recorrido y un número considerable de usuarios.

- **OGPL^[68]**: es un proyecto conjunto de colaboración entre los gobiernos de Estados Unidos, Canadá y la India para promover la apertura de datos públicos mediante la creación de una plataforma de código abierto. El proyecto incluye módulos para la gestión de los datos, la publicación de un catálogo web, la gestión de contenidos adicionales y la interacción con la comunidad.

Es de suma importancia entender que más allá de las características generales existentes en la mayoría de las herramientas mencionadas, se debe tener en consideración las regulaciones y estándares definidos a nivel nacional e internacional sobre los datos abiertos.

Análisis de beneficios del prototipo frente a otros portales de datos abiertos

Como se ha mencionado en capítulos anteriores, las herramientas disponibles para el análisis, gestión y disponibilización de datos abiertos, se ha visto reducida a un conjunto de empresas o comunidades que optimizan dichos productos. Esto genera conflictos cuando se encuentran limitaciones en sus prestaciones o que no cumplen con los requisitos pretendidos por algún equipo usuario de datos abiertos, lo que crea un replanteo en la definición de las características que el sistema de software debe abastecer y la decisión de llevar a cabo un desarrollo propio a medida de las necesidades.

CKAN y Junar permiten a sus usuarios administrar y gestionar los datos abiertos por medio de un panel de carga de información y pone a disposición el uso de APIs que automatizan y agilizan la interacción entre los portales y entes externos.

Sobre estas dos herramientas se ha realizado un detallado análisis técnico y funcional para determinar las vulnerabilidades que pueden surgir a la hora de disponibilizar los datos abiertos. También se ha realizado una comparativa con los resultados del relevamiento de algunos beneficios considerables a la hora de optar por un desarrollo propio para resolver la apertura de la información de la forma más adecuada y por lo tanto la necesidad de realizar el prototipo presentado en esta Tesina, ODIN.

Luego de analizar la herramienta CKAN, se obtuvieron las siguientes conclusiones:

Usabilidad:

- El panel de administración es poco intuitivo al momento de gestionar el contenido de la plataforma por el usuario final. El prototipo desarrollado cuenta con un panel de administración intuitivo y amigable para el usuario.
- La plataforma Frontend que presenta CKAN es escueta en la medida de usabilidad, estructura y simplicidad para el usuario final. Realizar un desarrollo complementario y luego integrarlo resulta engorroso y conlleva una carga notable de trabajo. El prototipo mejora notablemente la estética, haciéndola más agradable y simple de cara al ciudadano. Es un desarrollo independiente de la plataforma administrativa y del sistema de software del lado servidor.

- La subida de archivos en CKAN corresponde a la necesidad de scripts externos o un servidor de FTP. En el prototipo, la subida de archivos se puede realizar directamente desde la plataforma.
- El panel de administración de CKAN solo cuenta con un buscador general sobre el nombre de la entidad en cuestión. El panel de administración del prototipo cuenta con un buscador por múltiples campos.
- CKAN cuenta con una limitación en los roles de usuario permitidos en el panel de administración, mientras que en el prototipo presentado se incorpora el manejo de roles de usuario correspondientes a tres tipos diferentes: Superadministrador, Administrador e Invitado. Este manejo de roles permite al usuario Superadministrador crear usuarios con diferentes permisos sobre las secciones del panel administrativo de la plataforma. El rol de mayor interés es el Invitado, ya que tiene un acceso limitado sobre las secciones que presenta el panel de administración: sólo puede crear recursos y visualizaciones relacionadas a dichos recursos; además, podrá editar sólo aquellos creados por él. Estos recursos y/o visualizaciones no serán publicados hasta que no sean aprobados por un usuario Administrador o Superadministrador.
- La plataforma Frontend que presenta CKAN no presenta módulos de comentarios disponibles para los usuarios finales. El prototipo presenta un módulo de comentarios sobre los conjuntos de datos llevado a cabo mediante la herramienta Disqus.
- La plataforma Frontend que presenta CKAN no cuenta con estadísticas de descargas de los recursos y los conjuntos de datos. El prototipo pone a disposición dichas estadísticas como muestra de la interacción que presenta el portal de datos abiertos con los usuarios del mismo.
- CKAN no presenta un historial de actividades sobre los módulos correspondientes al panel de administración. El prototipo lleva adelante un control de logs sobre cada uno de los módulos del sistema con el fin de detallar la actividad que presenta permite ver las interacciones de los usuarios (especificando al usuario en cuestión), ya sean creaciones, ediciones o borrados de los módulos pertinentes.
- La plataforma Frontend que presenta CKAN pone a disposición una guía de datos correspondiente a cada conjunto de datos individualmente como un recurso más en su listado. El prototipo disponibiliza la guía de datos como una sección importante dentro del contenedor correspondiente al conjunto de datos y se diferencia del resto de los recursos. Además cuenta con una previsualización incorporada dentro del mismo portal. La guía de datos corresponde a información relevante sobre los datos disponibilizados con que cuenta el conjunto de datos en cuestión y es confeccionado, generalmente, por el equipo de datos abiertos de la institución.

Concepto del conjunto de datos:

- La herramienta CKAN no cuenta con un manejo de estados por lo que todo conjunto de datos y recurso creado desde el panel de administración se mostrará automáticamente en la plataforma Frontend correspondiente. El prototipo dispone de un manejo de estados tanto en los conjuntos de datos, como en los recursos y sus

visualizaciones. Esto permite al usuario de la plataforma la publicación o no de alguna de estas entidades; así mismo, la posibilidad de enviar a revisión o rechazar según corresponda. Los estados con los que cuenta el prototipo son: Publicado, En revisión, Borrador, Rechazado, Quitado de publicación.

- Una de las limitaciones más importantes encontradas en la herramienta CKAN es la falta de control sobre los archivos cargados en la plataforma como dato abierto. No permite búsquedas sobre el contenido de los archivos asociados a texto como suelen ser los formatos CSV, XLS, XLSX y JSON, por lo que solo se encuentra el archivo físico para su descarga. El prototipo genera un parseo de la información contenida en los archivos físicos que corresponden a los formatos CSV, XLS, XLSX y JSON y los almacena en una base de datos no relacional. De esta manera, los cuatro formatos son accesibles de una única forma respetando la estructura del formato JSON y permite acceder de forma ágil y realizar búsquedas sobre cada uno de los campos que contenga.
- La herramienta CKAN presenta únicamente dos fechas relevantes tanto en los recursos como en los conjuntos de datos y corresponden a la fecha de creación y de actualización del módulo en cuestión, mientras que el prototipo desarrollado contempla estas dos fechas y también las correspondientes a los estados por los que pueden transitar cada módulo. A su vez, es importante destacar que los recursos cuentan con una fecha adicional correspondiente a la fecha de relevamiento de la información.

Calidad de los datos:

- CKAN no cuenta con una sección dedicada a la carga de organizaciones correspondientes a la fuente del dato disponibilizado, esta asociación se hace mediante categorías o etiquetas. El prototipo pone a disposición una sección específica para la administración de las organizaciones a cargo de los datos disponibilizados en la plataforma, lo que permite mantener una normalización de las mismas en todo el ámbito del proyecto.
- La herramienta CKAN cuenta con un acotado número de atributos por defecto en cada uno de los módulos de su sistema, por lo cual agregar información estándar para dicho módulo debe realizarse a través de la opción de campos adicionales, lo que genera que el usuario encargado lleve un control sobre los datos adicionales para corresponder de la misma manera en todas las secciones correspondientes. El prototipo desarrollado cuenta con todos los campos necesarios en cada uno de los módulos correspondientes, lo que permite a su vez la búsqueda, filtrado y tratamiento especial por cada uno de ellos. Esta integración se debe a un análisis previo con equipos dedicados a datos abiertos para dar con los atributos necesarios y provechosos para su gestión.
- Si bien la herramienta CKAN cuenta con una sección dentro de la carga de los módulos de conjuntos de datos y de recursos donde se pueden agregar campos adicionales, tiene la limitación de que al agregar un nuevo conjunto de datos o un recurso, no mantiene estos campos. Por lo que resulta tedioso en el caso de que todos los conjuntos de datos o recursos deban ser cargados con dichos campos adicionales y cada vez deban ser creados. Además, se encuentra en una situación

en la que pueden no respetar la misma sintaxis y quedarán visualmente mal. El prototipo permite la definición de campos adicionales por defecto tanto para los conjuntos de datos como para los recursos y de esta manera agilizar la carga de los mismos a la hora de crear alguna de estas entidades.

- CKAN no permite la previsualización de archivos con formato PDF, mientras que el prototipo contempla la previsualización de este tipo de archivos tanto en la plataforma Frontend como en el panel de administración con el fin de evitar descargar el archivo para ver su contenido.
- Si bien CKAN cuenta con la posibilidad de asociar visualizaciones con los recursos correspondientes no cuenta con su previsualización desde el panel de administración. El prototipo contempla esta funcionalidad tanto para las visualizaciones de gráficos, mapas y archivos con formato PDF como fue detallado anteriormente.

Presentar la información:

- La herramienta CKAN cuenta con la posibilidad de asociar visualizaciones con los recursos correspondientes mediante las integraciones desarrolladas en dicha herramienta lo cual plantea una dependencia por parte del usuario con dicha integración. El prototipo permite asociar visualizaciones en forma de gráfico y mapa no sólo embebiendo un iframe desde un link externo, sino también generarlos desde la misma plataforma a partir del parseo mencionado anteriormente de los archivos cargados con los formatos CSV, XLS, XLSX y JSON.
- La herramienta CKAN permite la descarga de los recursos en su formato correspondiente al cual fue subido desde el panel de administración. El prototipo cuenta con el adicional de permitir la descarga de archivos correspondientes a CSV, XLS y XLSX en cualquiera de sus formatos. Esto quiere decir que un archivo CSV, no solo podrá ser descargado en formato CSV, sino que podrá ser descargado en XLS o XLSX, y así para cada una de las combinaciones entre los tres formatos de archivo.

Mejoras técnicas del prototipo propuesto para la tesina:

- Está diseñado con un conjunto de tres proyectos independientes: cuenta con una API que interactúa con el panel de administración y con la plataforma Frontend correspondiente y es también, el que se encarga de la persistencia y comunicación con las bases de datos dispuestas. Además, esta arquitectura permite futuras plataformas orientadas al cliente como pueden ser aplicaciones móviles o web, utilizando la misma API.
- Posibilita búsquedas tanto en los metadatos de los recursos como también dentro de los archivos CSV, XLS Y XLSX.
- Almacena el contenido de los archivos gestionados en el panel de administración en una base de datos para su posterior uso y consulta.

- Posibilita la carga de recursos mediante servicios web de tipo REST y SOAP con el fin de agilizar la gestión de la información dispuesta en la plataforma. A su vez, permite la interacción de infraestructura sensorial con la plataforma para generar conjuntos de datos en tiempo real.
- Cuenta con un importador de información desde la herramienta CKAN para permitir a las entidades que cuentan con esta herramienta la migración de su información.
- Es un proyecto de código abierto y gratuito lo que conlleva libertades tanto legales como económicas para las entidades que lo utilicen para su apertura de datos abiertos. El prototipo se encuentra bajo licencia GNU General Public License v3.0^[69] que permite su modificación y distribución sin restricciones ni fines de lucro.

Luego de analizar la herramienta Junar, se ha realizado un informe sobre las siguientes características:

Usabilidad:

- La plataforma Frontend que presenta Junar no cuenta con estadísticas de descargas de los recursos y los conjuntos de datos. El prototipo desarrollado pone a disposición dichas estadísticas como muestra de la interacción que presenta el portal de datos abiertos con los usuarios del mismo.
- La plataforma Frontend que presenta Junar no cuenta con una guía de datos que explique el contenido de la información que se muestra. El prototipo disponibiliza la guía de datos como una sección importante dentro del contenedor correspondiente al conjunto de datos. Además cuenta con una previsualización incorporada dentro del mismo portal. La guía de datos corresponde a información relevante sobre los datos disponibilizados con que cuenta el conjunto de datos en cuestión y es confeccionado, generalmente, por el equipo de datos abiertos de la institución.

Concepto del conjunto de datos:

- Junar dispone el concepto dataset como un recurso que puede ser creado a través de un archivo físico o un servicio web y luego puede ser agrupado por diferentes categorías. El prototipo define el concepto dataset como un conjunto de datos, estos datos se los conoce como recursos. A su vez, el conjunto de datos puede ser asociado a más de una categoría logrando con ello un filtro más a la hora de buscar los datos dentro de la plataforma.
- Junar cuenta con dos estados para los módulos de su plataforma y estos estados corresponden a Borrador y Publicado. El prototipo, como se mencionó anteriormente, cuenta con cinco estados por los que pueden pasar cada uno de los módulos de la plataforma. Esta característica tiene gran utilidad cuando se cuenta con un usuario Invitado, ya que dicho usuario puede crear recursos o visualizaciones en estado Borrador y enviar a Revisión una vez que haya terminado de relevar toda la información necesaria para la generación de dicho módulo.

Calidad de los datos:

- Junar no cuenta con una sección dedicada a la carga de organizaciones correspondientes a la fuente del dato disponibilizado, esta asociación se hace mediante categorías o etiquetas. El prototipo pone a disposición una sección específica para la administración de las organizaciones a cargo de los datos disponibilizados en la plataforma y esto permite mantener una normalización de las mismas en todo el ámbito del proyecto.

Aspecto legal y licencia:

- La herramienta Junar es un desarrollo propietario y las modificaciones pertinentes para satisfacer las necesidades de quien utilice esta herramienta es responsabilidad de la empresa Junar. El prototipo desarrollado es un proyecto de software libre y de código abierto, por lo cual las modificaciones pertinentes para satisfacer las necesidades de apertura de la información podrá ser llevada a cabo por el equipo de desarrollo de la entidad que lo utilice.
- Junar posee todos los derechos sobre el software, el código, las bases de datos y otras aplicaciones de servicios, permitiendo al cliente conservar los derechos sobre los datos. El prototipo es de código abierto y con una licencia GNU General Public License v3.0 que permite su modificación y distribución sin restricciones ni fines de lucro, los derechos recaen sobre quien utilice dicho producto, tanto de las tecnologías utilizadas como de los datos almacenados.
- La herramienta Junar al ser un desarrollo propietario es costoso para la economía de la entidad que la utilice. Esto puede presentar limitaciones para gobiernos emergentes que no puedan hacer grandes erogaciones de dinero debido a sus ajustadas arcas financieras. El prototipo es completamente gratuito y sus eventuales modificaciones dependerá exclusivamente de contar con un equipo de desarrollo interno en la entidad que lo utilice.

A continuación se presenta una tabla comparativa de las características evaluadas sobre las tres herramientas analizadas:

Características		CKAN	Junar	ODIN
Usabilidad	Administrador intuitivo y amigable	No	Si	Si
	Catálogo de datos como desarrollo independiente	No	Si	Si
	Carga de archivos directo desde el administrador	No	Si	Si
	Carga de recursos mediante servicio web	No	Si	Si
	Buscador con campos múltiples	No	No	Si
	Múltiples roles de usuario	No	Si	Si
	Módulo de comentarios	No	No	Si
	Estadísticas de descarga	No	No	Si
	Historial de actividades en el administrador	No	Si	Si
	Guía de datos	Si	No	Si
Concepto del conjunto de datos	Manejo de estados múltiples	No	No	Si
	Búsqueda sobre contenido de archivos físicos subidos a la plataforma	No	No	Si
	Fechas correspondientes a transiciones de estados	No	No	Si
Calidad de los datos	Módulo de entidades relacionales a los conjuntos de datos	No	No	Si
	Limitación de atributos por entidad	Si	Si	No
	Previsualización de archivos PDF	No	No	Si
	Previsualización de visualizaciones en el administrador	No	Si	Si

Presentación de la información	Visualizaciones de los recursos mediante links embebidos	No	Si	Si
	Gestión personalizada de visualizaciones	No	Si	No
	Selección de formato de descarga múltiple	No	No	Si
Aspectos legales y licencia	Software libre	Si	No	Si
	Código abierto	Si	No	Si
	Derechos sobre el software y datos	Comunidad	Propietaria	Comunidad
	Licencia	GNU GPL v3.0	Propietaria	GNU GPL v3.0
Otros	Importador de datos desde CKAN	Si	No	Si
	Proyecto en Github	https://github.com/ckan/ckan	https://github.com/Junar	https://github.com/odin-unlp

Resumen

Es de gran importancia realizar un análisis exhaustivo del proyecto para determinar cuáles serán las tecnologías, arquitecturas o herramientas que se necesitarán para llevarlo adelante y cumplir con los requisitos pretendidos. La gran cantidad de tecnologías presenta variedad de elección, pero a su vez genera la necesidad de evaluar los beneficios de utilizar unas u otras. Así mismo, optar por un desarrollo propio tiene como consecuencia una mayor flexibilidad en el modelado del sistema con lo cual se obtendrá un producto a la medida de las necesidades esperadas. Además, tomando las decisiones correctas, permite la escalabilidad de la herramienta en cuestión y la extensibilidad por medio de la integración con diferentes desarrollos alternativos desarrollados por terceros para optimizar o gestionar diferentes módulos. Por dicho motivo, si bien existen plataformas comerciales y de software libre para disponibilizar los datos, se han presentado muchas limitaciones que redundan en un prototipo que se adecue a las necesidades de cualquier organización.

Referencias

- [1] <https://medium.freecodecamp.org/what-is-an-api-in-english-please-b880a3214a82>, fecha de consulta: 3 de junio 2018.
- [2] <https://bbvaopen4u.com/es/actualidad/api-rest-que-es-y-cuales-son-sus-ventajas-en-el-desarrollo-de-proyectos>, fecha de consulta: 3 de junio 2018.
- [3] <https://blogs.mulesoft.com/biz/tech-ramblings-biz/what-are-apis-how-do-apis-work/>, fecha de consulta: 3 de junio 2018.
- [4] <https://hipertextual.com/archivo/2014/05/que-es-api/>, fecha de consulta: 3 de junio 2018.
- [5] https://es.wikipedia.org/wiki/Transferencia_de_Estado_Representacional, fecha de consulta: 3 de junio 2018.
- [6] <https://searchmicroservices.techtarget.com/definicion/REST-representational-state-transfer>, fecha de consulta: 3 de junio 2018.
- [7] https://www.ics.uci.edu/~fielding/pubs/dissertation/rest_arch_style.htm, fecha de consulta: 3 de junio 2018.
- [8] <http://microservices.io/patterns/apigateway.html>, fecha de consulta: 3 de junio 2018.
- [9] <https://www.nginx.com/blog/building-microservices-using-an-api-gateway/>, fecha de consulta: 3 de junio 2018.
- [10] <http://getkong.org/>, fecha de consulta: 3 de junio 2018.
- [11] <https://aws.amazon.com/es/relational-database/>, fecha de consulta: 3 de junio 2018.
- [12] <https://www.mysql.com>, fecha de consulta: 15 de agosto 2018.
- [13] <https://mariadb.org>, fecha de consulta: 15 de agosto 2018.
- [14] <https://www.microsoft.com/es-es/sql-server>, fecha de consulta: 15 de agosto 2018.
- [15] <https://www.oracle.com/database/index.html>, fecha de consulta: 15 de agosto 2018.
- [16] <https://www.postgresql.org/about/>, fecha de consulta: 3 de junio 2018.
- [17] <http://postgis.net/>, fecha de consulta: 15 de agosto 2018.
- [18] <https://insights.stackoverflow.com/survey/2018/>, fecha de consulta: 3 de junio 2018.
- [19] <https://aws.amazon.com/es/nosql/>, fecha de consulta: 3 de junio 2018.
- [20] <https://www.mongodb.com/nosql-explained>, fecha de consulta: 3 de junio 2018.
- [21] <https://www.mongodb.com/company?jmp=footer>, fecha de consulta: 3 de junio 2018.
- [22] <https://developer.mozilla.org/es/docs/Glossary/JavaScript>, fecha de consulta: 4 de junio 2018.
- [23] http://librosweb.es/libro/javascript/capitulo_1/breve_historia.html, fecha de consulta: 4 de junio 2018.
- [24] <http://www.utel.edu.mx/blog/menu-profesional/historia-del-lenguaje-javascript/>, fecha de consulta: 4 de junio 2018.
- [25] <https://nodejs.org/en/>, fecha de consulta: 5 de junio 2018.
- [26] <https://stackoverflow.com>, fecha de consulta: 9 de junio 2018.
- [27] <https://insights.stackoverflow.com/survey/2018/#technology>, fecha de consulta: 5 de junio 2018.
- [28] <https://www.npmjs.com>, fecha de consulta: 5 de junio 2018.
- [29] <https://sailsjs.com/about>, fecha de consulta: 9 de junio 2018.
- [30] <https://github.com/balderdashy/sails>, fecha de consulta: 9 de junio 2018.
- [31] <https://angularjs.org/>, fecha de consulta: 4 de junio 2018.
- [32] <https://www.sebastian-gomez.com/desarrollo-web/que-es-angularjs-parte-1/>, fecha de consulta: 4 de junio 2018.
- [33] <https://carlosazaustre.com/empezando-con-angular-js/>, fecha de consulta: 4 de junio 2018.
- [34] <https://www.google.com/>, fecha de consulta: 4 de junio 2018.
- [35] <http://es6-features.org/>, fecha de consulta: 6 de junio 2018.
- [36] <https://www.dartlang.org/>, fecha de consulta: 6 de junio 2018.
- [37] <https://coffeescript.org/>, fecha de consulta: 6 de junio 2018.

- [38] <https://www.polymer-project.org/>, fecha de consulta: 6 de junio 2018.
- [39] <https://x-tag.github.io/>, fecha de consulta: 6 de junio 2018.
- [40] <http://backbonejs.org/>, fecha de consulta: 6 de junio 2018.
- [41] <https://jquery.com/>, fecha de consulta: 6 de junio 2018.
- [42] <https://underscorejs.org/>, fecha de consulta: 6 de junio 2018.
- [43] <https://jasmine.github.io/>, fecha de consulta: 6 de junio 2018.
- [44] <https://karma-runner.github.io/2.0/index.html>, fecha de consulta: 6 de junio 2018.
- [45] <https://gulpjs.com/>, fecha de consulta: 7 de junio 2018.
- [46] <https://platzi.com/blog/automatizacion-gulp-js/>, fecha de consulta: 7 de junio 2018.
- [47] <http://blog.koalite.com/2015/06/grunt-o-gulp-que-uso/>, fecha de consulta: 7 de junio 2018.
- [48] <https://www.adictosaltrabajo.com/tutoriales/primeros-pasos-con-gulp/>, fecha de consulta: 7 de junio 2018.
- [49] <https://www.emberjs.com/>, fecha de consulta: 7 de junio 2018.
- [50] <https://reactjs.org/>, fecha de consulta: 7 de junio 2018.
- [51] <https://www.java.com/>, fecha de consulta: 7 de junio 2018.
- [52] <https://maven.apache.org/>, fecha de consulta: 7 de junio 2018.
- [53] <http://groovy-lang.org/>, fecha de consulta: 7 de junio 2018.
- [54] <https://gradle.org/>, fecha de consulta: 7 de junio 2018.
- [55] <https://gruntjs.com/>, fecha de consulta: 7 de junio 2018.
- [56] <https://nodejs.org/api/stream.html>, fecha de consulta: 7 de junio 2018.
- [57] <https://blog.kezmo.com/qu%C3%A9-son-las-metodolog%C3%ADas-%C3%A1giles-y-por-qu%C3%A9-debes-implementarlas-en-tu-organizaci%C3%B3n-484a510e5b0>, fecha de consulta: 8 de junio 2018.
- [58] <https://blog.conectart.com/metodologias-agiles/>, fecha de consulta: 8 de junio 2018.
- [59] <http://datos.gob.es/sites/default/files/informe-herramientas-publicacion.pdf>, fecha de consulta: 10 de julio 2018.
- [60] <http://linkeddata.org/>, fecha de consulta: 12 de julio 2018.
- [61] <https://www.w3.org/standards/semanticweb/>, fecha de consulta: 12 de julio 2018.
- [62] <https://www.w3.org/TR/ldp/>, fecha de consulta: 12 de julio 2018.
- [63] <https://www.w3.org/TR/sparql11-overview/>, fecha de consulta: 12 de julio 2018.
- [64] <https://www.drupal.org/>, fecha de consulta: 15 de julio 2018.
- [65] <https://github.com/datagovuk/dgu-vagrant-puppet/blob/togo/README.md>, fecha de consulta: 15 de julio 2018.
- [66] <https://datapress.com/>, fecha de consulta: 15 de julio 2018.
- [67] <https://www.opendatasoft.com/>, fecha de consulta: 15 de julio 2018.
- [68] <http://ogpl.github.io/index-en.html>, fecha de consulta: 15 de julio 2018.
- [69] <https://www.gnu.org/licenses/gpl-3.0.en.html>, fecha de consulta: 14 de junio 2018.

Prototipo

Introducción

El avance de las tecnologías open source en el mercado informático, permite que a la hora de realizar un desarrollo se cuente con una gran variedad de tecnologías y herramientas disponibles. Esto conlleva a la realización de un análisis y una evaluación comparativa entre las distintas herramientas, para así elegir las que resulten más adecuadas para acercarse a la solución esperada. Por otro lado, este incremento en el caudal de tecnologías generó que cada una se especializara en temáticas particulares, por lo cual, para llevar a cabo el desarrollo de un sistema robusto y escalable se necesitará la colaboración de varias herramientas que, trabajando en conjunto, permitan de manera más eficiente la resolución del problema.

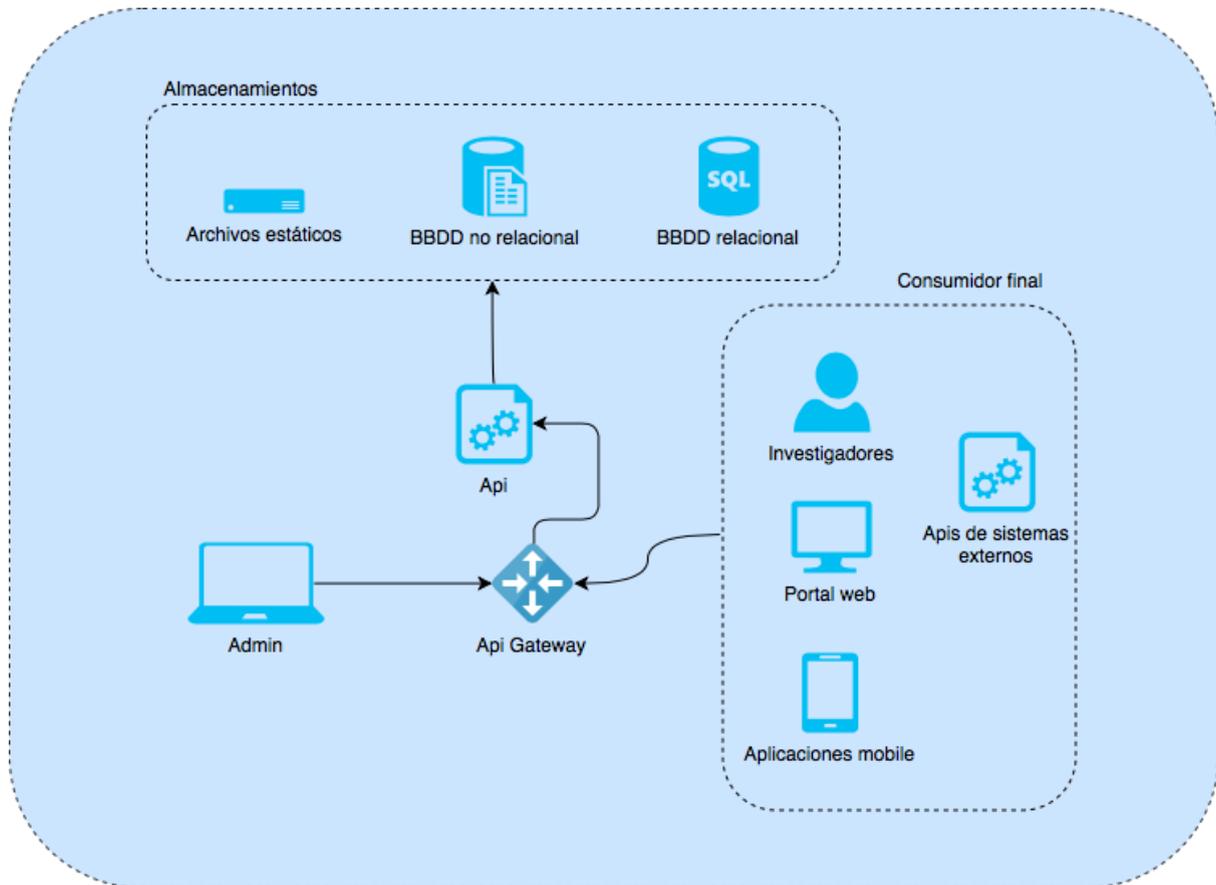
En este capítulo se detallará brevemente el ecosistema logrado por las herramientas que fueron descritas en el capítulo anterior para el desarrollo del prototipo de esta Tesina. Además, se mencionarán los pasos a seguir en la apertura de datos de la Universidad Nacional de La Plata y la información reclutada y generada para las pruebas necesarias del prototipo.

Cabe destacar que el plan de acción de apertura de la información es el pilar fundamental de toda política e iniciativa de datos abiertos^[1]. Es a través del cual:

- Se plantean los objetivos.
- Se realizan los análisis pertinentes sobre el valor de la información de la entidad de interés.
- Se relevan los posibles actores del ecosistema de datos abiertos.
- Se obtienen los datos a ser disponibilizados.
- Se realizan los procesos de normalización y estandarización pertinentes.
- Se determinan la/las licencias mediante las cuales serán disponibilizados los datos.
- Se realiza la apertura propiamente dicha.
- Se establecen los medios por los cuales se fomentará la utilización de los datos vinculados y específicos de la entidad.

Por último, se pondrán a disposición los resultados de las pruebas de uso realizadas con el objetivo de evaluar el prototipo mediante la ejecución de tareas específicas sobre perfiles expertos y no expertos, tanto en la gestión de los conjuntos de datos como en la interacción y consumo de la información a través del catálogo de datos.

Conceptos generales del prototipo



Ecosistema de herramientas que interactúan en Odin.

El sistema cuenta con una mezcla heterogénea de herramientas que trabajando en conjunto dentro de un mismo ecosistema cumplen su propósito de la mejor manera. Cada herramienta que se utilizó cumple un rol puntual en el proyecto. A continuación se va a describir cómo estas tecnologías previamente enunciadas interactúan entre sí para generar el sistema final.

En la imagen donde se detalla de forma figurativa el ecosistema e interacciones de Odin, podemos apreciar que en el centro de la misma se encuentra la API de consumo de datos. Esta API es un desarrollo REST, la cual permite la interacción entre los distintos motores de base de datos y el cliente final; es decir, cuando un cliente, ya sea mediante una página web o un desarrollo tercero, realiza una petición a la API, ésta hará lo necesario para brindar la información solicitada mediante políticas de conexión y seguridad con sus respectivas bases de datos configuradas. Asimismo, en la API podemos encontrar el manejo de autenticación y autorización a nivel sistema, guardado de archivos físicos, descarga de archivos en distintos formatos, generación de mapas, gráficos y filtros en los recursos, etc.

Muy cercano a la API de consumo de datos, podemos encontrar el gateway, que es el encargado de la seguridad y autenticación principal de la API, a nivel servidor. La herramienta utilizada es Kong, que a su vez utiliza Nginx como servidor de base. El valor agregado por esta herramienta, además de seguridad y autenticación, es entre otras cosas, el rate limiting, cors, métodos HTTP aceptados, fácilmente configurable y flexible.

Para el almacenamiento de datos se utilizan tres formatos diferentes:

- Por un lado, para guardar toda la metadata generada en la plataforma se usa una base de datos relacional, en nuestro caso la elección fue PostgreSQL, pero podría utilizarse cualquier otra base de datos relacional. En esta base de datos se guardan todos los datos que se generan como información adicional a los archivos subidos, usuarios, categorías y todo lo relacionado a la plataforma, como también las configuraciones del API Gateway (Kong).
- Por otro lado se utiliza una base de datos no relacional. En este caso la herramienta seleccionada fue MongoDB, donde se guardan los archivos que pueden ser transformados al formato JSON. Cuando un usuario que utiliza el administrador sube un archivo de tipo CSV, XLSX, XLS, JSON; la API lo va a recibir, sea el formato que sea, lo va a transformar a JSON y lo va a guardar en una colección de MongoDB, donde el nombre de la colección es el ID del recurso subido y el nombre de la Base es el ID del dataset contenedor; siendo así que un dataset que tenga muchos archivos subidos en los formatos mencionados va a poseer esa cantidad de colecciones en su base de datos. Tener los datos guardados de esta manera, permite una fácil y rápida previsualización del archivo en sí como también la posibilidad de transformar el JSON a distintos formatos de salida.
- Por último, para el guardado de los archivos estáticos se utiliza el File System. Cada vez que un usuario genere un nuevo recurso desde el administrador, éste archivo subido se guardará automáticamente en el File System del servidor. También se utiliza esta forma de almacenamiento para el armado de un archivo comprimido (zip) de los datasets, el cual posee todos los recursos que componen el mismo, sumado a un archivo de resumen con metadata del dataset.

Otro elemento del ecosistema planteado es el administrador, donde mediante un usuario y una contraseña se hará administración y gestión de los recursos, datasets, usuarios, configuraciones y todos los elementos que componen la herramienta. Es la única entrada desde donde se carga la información al sistema. Con los distintos roles de usuarios, es simple limitar las acciones que puedan tomar los distintos perfiles, permitiendo que un administrador general deba aceptar los datos subidos por otros.

Por último encontramos las aplicaciones de cara a los usuarios finales. Dentro del proyecto se incluye un desarrollo de una plataforma web, donde se muestra toda la información cargada desde el administrador mediante la interacción con la API. Sin embargo, la API al ser abierta cuenta con la posibilidad de generar aplicaciones móviles y conectar con otras APIs u otros sistemas utilizando los datos cargados originalmente.

Análisis de la apertura de datos

La apertura de la información es un pilar para la colaboración y participación ciudadana independientemente del área que la realice y, por consiguiente, un generador de interacción entre los diversos actores de una comunidad; esto aporta a la transparencia y a la realización de nuevos e innovadores productos.

Como se mencionó en el primer capítulo de este documento, para llevar a cabo la apertura de la información correspondiente a la Universidad Nacional de La Plata se ha seguido un detallado procedimiento de análisis, identificación y selección de los datos a disponibilizar. Una vez finalizadas estas etapas, se realizó la apertura de datos a través del prototipo desarrollado para esta Tesina.

El procedimiento para llegar al objetivo fue el siguiente:

- **Identificar los datos:** se realizó una búsqueda de datos dentro de las plataformas que corresponden al ámbito de la Universidad Nacional de La Plata con el fin de elaborar un listado con la información encontrada. Seguido a esto se llevó adelante la tarea de identificar a qué área correspondía la información y analizar qué tipo de visualización podía surgir de cada conjunto de datos. Luego se evaluaron diferentes mecanismos de normalización de datos para mantener una coherencia en la estructura para que a la hora de ser consumida por terceros fuera de forma legible y clara.

La información obtenida fue la siguiente:

- Correspondencias edilicias ligadas a la universidad.
- Becas administradas por la universidad.
- Resultados de las elecciones de los centros de estudiantes de cada una de las facultades correspondientes a la universidad.

Además, se confeccionó un documento adicional con información relacionada a los estudiantes inscriptos en cada una de las carreras que dicta la Universidad Nacional de La Plata. El mismo tiene como objetivo demostrar el impacto de la apertura de la información y los diferentes beneficios que tiene la utilización de una herramienta dedicada a este fin. La información que presenta el documento fue elaborada exclusivamente para esta Tesina y no contiene datos reales.

Es importante mencionar que el análisis realizado en esta etapa tiene como punto de partida información encontrada en los portales de datos abiertos de las distintas universidades del mundo mencionadas en capítulos anteriores. De allí se desprende la evaluación de las plataformas de la Universidad Nacional de La Plata y la consecuente identificación de los datos.

- **Seleccionar los datos:** tras identificar los datos a utilizar, se analizó el impacto de la apertura de los mismos en base a distintas variables como: cuáles serían los actores y las áreas involucradas, qué costo tendría la confección continua de esta información, y de qué manera se verían beneficiados la universidad y los consumidores finales de los datos.

La selección de esta información se debe a la facilidad de su confección y mantenimiento por parte de un equipo de datos abiertos con el que puede contar la institución. Por ejemplo, los datos relacionados con las becas corresponden tanto a las políticas internas, como a las establecidas por la institución superior y las interacciones con diferentes universidades del mundo; o los datos edilicios ligados a la universidad, corresponden a las instituciones administrativas y de recreación, donde los estudiantes pueden realizar trámites o eventuales visitas.

Es bueno tener presente que este tipo de información figura por defecto en cada uno de los portales evaluados en secciones anteriores correspondientes al ámbito académico, por lo que se decidió incluirlos dentro de la plataforma de datos abiertos de la Universidad Nacional

de La Plata.

Es importante saber que dentro de los beneficios de la disponibilización de estos datos se encuentra la confección de mapas. Los mismos ponen a disposición de forma gráfica las ubicaciones de las instituciones relevadas junto con información útil, dentro de una plataforma que centraliza la información y agiliza la interacción con quienes la requieran. Además, fomenta la creación de futuros proyectos a partir de los datos, ya que abre la puerta al desarrollo de soluciones como aplicaciones móviles, por ejemplo para que los estudiantes concentren la información sin la necesidad de estar realizando una búsqueda alternativa en internet.

	A	B	C	D	E	F	G	H	I	J
	Nombre	Decano	Ubicación	Localidad	Código Postal	Teléfonos	Mail	Sitio	Latitud	Longitud
1	Facultad de Arquitectura y Urbanismo	Arq. Fernando Gandolfi	Calle 47 y 117	La Plata	1900	(0221) 423-6587	fau@fau.unlp.edu	http://www.fau.unlp.edu	-34.904706	-57.942441
2	Facultad de Bellas Artes	Dr. Daniel Belinche	Diagonal 78 680	La Plata	1900	(0221) 423-6598	academica@fba.unlp.edu	http://www.fba.unlp.edu	-34.921954	-57.941611
3	Facultad de Ciencias Agrarias y Forestales	Ing. Ricardo Andreau	Av. 60 y 119	La Plata	1900	(0221) 423-6758	saa@agro.unlp.edu	http://www.agro.unlp.edu	-34.910465	-57.929024
4	Facultad de Ciencias Astronómicas y Geofísicas	Lic. Raúl Perdomo	Paseo del Bosque	La Plata	1900	(0221) 423-6593	extension@fcaglp.unlp.edu	http://www.fcaglp.unlp.edu	-34.908312	-57.933207
5	Facultad de Ciencias Económicas	Mg. Eduardo Andrés De Giusti	Calle 6 e/ 47 y 48	La Plata	1900	(0221) 423-6771	sec-academica@conce.unlp.edu	http://www.conce.unlp.edu	-34.911679	-57.950507
6	Facultad de Ciencias Exactas	Dr. Mauricio Erben	Calle 47 y 115	La Plata	1900	(0221) 422-6977	academic@deca.unlp.edu	http://www.exactas.unlp.edu	-34.9063	-57.944189
7	Facultad de Ciencias Jurídicas y Sociales	Miguel Ángel Berri	Calle 48 e/ 6 y 7	La Plata	1900	(0221) 644-2070	info@jursoc.unlp.edu	http://www.jursoc.unlp.edu	-34.913223	-57.95009
8	Facultad de Ciencias Médicas	Dr. Juan Ángel Basualdo Farjat	Calle 60 y 120	La Plata	1900	(0221) 424-1596	alumnado@med.unlp.edu	http://www.med.unlp.edu	-34.908883	-57.928166
9	Facultad de Ciencias Naturales y Museo	Dr. Ricardo Etcheverry	Calle 60 y 122	La Plata	1900	(0221) 423-6773	fcnym@musseo.unlp.edu	http://www.fcnym.unlp.edu	-34.908498	-57.925863
10	Facultad de Ciencias Veterinarias	Dr. Marcelo Pecoraro	Calle 60 y 118	La Plata	1900	(0221) 423-6663	prensaydifusion@fcv.unlp.edu	http://www.fcv.unlp.edu	-34.911355	-57.930451
11	Facultad de Humanidades y Ciencias de la Educación	E. Ana Julia Ramírez	Calle 51 e/ 124 y 125	Ensenada	1925	(0221) 423-0128	sae@fahce.unlp.edu	http://www.fahce.unlp.edu	-34.900225	-57.931569
12	Facultad de Informática	Lic. Patricia Pesado	Calle 50 y 120	La Plata	1900	(0221) 427-7270	difusion@info.unlp.edu	http://www.info.unlp.edu	-34.903386	-57.937262
13	Facultad de Ingeniería	Ing. Horacio Frene	Calle 1 y 47	La Plata	1900	(0221) 425-8911	secretaria@academica.unlp.edu	http://www.ing.unlp.edu	-34.90739	-57.944797
14	Facultad de Odontología	Dr. Gabriel Eduardo Lazo	Calle 1 y 50	La Plata	1900	(0221) 423-6775		http://www.folp.unlp.edu	-34.908891	-57.942665
15	Facultad de Periodismo y Comunicación Social	Andrea Varela	Diagonal 113 y 63	La Plata	1900	(0221) 425-0133	decanato@perio.unlp.edu	http://www.perio.unlp.edu	-34.912263	-57.926549
16	Facultad de Psicología	Lic. Xavier Oñativía	Calle 51 e/ 122 y 123	Ensenada	1925	(0221) 482-4415	comunicacion@psico.unlp.edu	http://www.psicologia.unlp.edu	-34.900191	-57.932853
17	Facultad de Trabajo Social	Dr. Néstor Artífano	Calle 9 y 63	La Plata	1900	(0221) 452-5317	comunicacion@trabajosocial.unlp.edu	http://www.trabajosocial.unlp.edu	-34.925095	-57.940504
18	Escuela Universitaria de Recursos Humanos Nery Orlando Fures		Calle 12 y 161	Berisso	1923	(0221) 644-7281	eurhtes@med.unlp.edu	http://www.eurhtes.unlp.edu	-34.875539	-57.886438

Muestreo de datos de las instituciones ligadas a la Universidad Nacional de La Plata.

Otro beneficio que surge tras la disponibilización de los datos es en cuanto a las becas, ya que mediante un análisis de los mismos se podrían mejorar las interacciones de los estudiantes con las oportunidades que administra la universidad. Hoy en día la información se encuentra limitada en una sección muy estructurada de la web principal de la UNLP, sin embargo, al abrir los datos, habría más visibilidad de los mismos y mediante la aplicación de ciertos filtros sobre diferentes variables como, instituciones que estimulan las becas, países que se encuentran involucrado, las remuneraciones estimadas o los plazos comprendidos, los interesados podrían tomar decisiones en base a información accesible, fácil de interpretar y confiable.

	A	B	C	D	E	F	G	H	I
1	Nombre	Descripción	Destino	Ubicación de cor	Teléfono	Mail	Sitio	Fecha Inicio	Fecha Fin
2	Centro Educativo Canadiense	Su objetivo es pc	Canadá	Av. de Mayo 776	(54 11) 4345-455	info@estudiare	http://www.estudiarecanada.org/		
3	Programa de becas en la Republica Popular China	La Oficina Cultur	China				https://www.arge	01/02/2018	01/03/2018
4	Programa de becas Saint Exupery (Francia)		Francia				https://www.argentina.gob.ar/educ		30/04/2018
5	Programa de becas del Gobierno Húngaro	Con el objetivo d	Hungría				https://www.argentina.gob.ar/notici		19/03/2018
6	Becas de Movilidad entre Universidades Andaluzas e Iberoamericanas	Este programa fi	España				http://www.auij.org/es/novedades-	20/09/2018	
7	Becas para cursar Másteres Oficiales en la Universidad de Cantabria	Dirigido a: latino	España				http://www.auij.org/es/becas-auij/1516		15/04/2018
8	Becas para cursar Másteres Universitarios en la Universidad Pública de	Dirigido a: estudi	España				http://www.auij.org/es/becas-auij/1514		30/04/2018
9	Becas para cursar Másteres Oficiales en la Universidade de Aveiro - Port	Dirigido a: estudi	Portugal				http://www.auij.org/es/becas-auij/1512		27/04/2018
10	Programa de becas Fulbright (Estados Unidos)	Su objetivo es pr	Estados Unidos	Viamonte 1653	(54 11) 4814-356	info@fulbright.co	http://fulbright.edu.ar/becas/para-g		19/04/2018
11	Centro Educativo Canadiense	Su objetivo es pc	Canadá	Av. de Mayo 776	Tel. (54 11) 4345	info@estudiare	http://estudiarecanada.org/%A0		
12	Nueva Zelanda	Becas de posgra Ser ciudadano d Aceptar regresar Cumplir con los r	Nueva Zelanda				https://www.arge	15/02/2018	28/03/2018
13	Fundación Carolina	Promueve la coo	España				http://www.fundacioncarolina.es		

Muestreo de datos de las becas administradas por la Universidad Nacional de La Plata.

La información concentrada en el documento adicional pretende demostrar la explotación de los datos orientados a big data, ya que a través del análisis concreto de esta información se pueden hacer, por ejemplo, estimaciones sobre las procedencias y edades de los aspirantes; cuáles son las carreras de la universidad más solicitadas; o qué margen existe de igualdad de género en la concurrencia a la educación universitaria. Esta explotación de la información permite fortalecer políticas con las ciudades de donde provienen mayor cantidad de aspirantes, fomentar la igualdad en carreras que presentan una disparidad, o establecer parámetros sobre motivos de inserción o deserción de los aspirantes a las carreras de grado que imparte la universidad.

	A	B	C	D	E	F	G	H	I
1	edad	sexo	pais	ciudad	escuela	facultad	carrera	latitud	longitud
2	25	Femenino	Chile	Posadas	ESC. PRIMARIA	Facultad de Cs. Económicas	Tec. en Cooperativismo	-31	-69
3	20	Femenino	Argentina	Bahía Blanca	INST. SUPERIO	Facultad de Cs. Jurídicas y Sociales	Tecnicultura Superior Universitaria de Martillero Público	-42	-68
4	26	Masculino	Argentina	Ciudad de Salta	CENOF D.E. 06	Facultad de Cs. Jurídicas y Sociales	Procurador	-30	-65
5	19	Masculino	Argentina	Bahía Blanca	INST. TOMAS DI	Facultad de Cs. Económicas	Lic. en Administración	-43	-66
6	21	Femenino	Argentina	Guaymallén	INST. DE NATAC	Facultad de Cs. Jurídicas y Sociales	Procurador	-36	-62
7	27	Masculino	Venezuela	Guaymallén	INST. ORAL MOI	Facultad de Arquitectura y Urbanismo	Arquitectura	-46	-70
8	25	Masculino	Colombia	La Plata	CENTRO EDUC.	Facultad de Cs. Exactas	Lic. en Ciencia y Tecnología de los Alimentos	-31	-61
9	25	Femenino	Argentina	Paraná	UGEE N° 01 UN	Facultad de Cs. Jurídicas y Sociales	Abogacía	-31	-60
10	32	Masculino	Colombia	Quilmes	CENTRO EDUC.	Facultad de Cs. Exactas	Lic. en Química	-46	-60
11	24	Masculino	Argentina	Pilar	INST. 9 DE JULI	Facultad de Cs. Agrarias y Forestales	Ingeniería Agronómica	-26	-70
12	26	Masculino	Venezuela	Lanús	UGEE N° 01 UN	Facultad de Bellas Artes	Lic. y/o Prof. en Historia del Arte orientación Artes Visua	-24	-67
13	23	Masculino	Argentina	San Isidro	INST. 9 DE JULI	Facultad de Cs. Exactas	Lic. en Óptica Ocular y Optometría	-30	-65
14	25	Masculino	Paraguay	Ciudad de Santi	CENOF CENTRI	Facultad de Cs. Astronómicas y Geofísicas	Licenciatura en Meteorología y Ciencias de la Atmosfer	-43	-57
15	22	Masculino	Argentina	Buenos Aires	INST. PRIVADO	Facultad de Cs. Exactas	Lic. en Biotecnología y Biología Molecular	-41	-62
16	31	Femenino	Argentina	Merlo	INST. TOMAS DI	Facultad de Bellas Artes	Lic. y/o Prof. en Diseño Multimedial (orientación: Audiov	-49	-58
17	25	Masculino	Chile	Paraná	INST. SUPERIO	Facultad de Cs. Jurídicas y Sociales	Técnico en Gestión de Recursos para Instituciones Unih	-26	-63
18	25	Femenino	Argentina	Bahía Blanca	ESC. PRIMARIA	Facultad de Bellas Artes	Lic. y/o Prof. en Música orientación Dirección Orquestal	-23	-60
19	31	Masculino	Argentina	Resistencia	INST. VOCACIO	Facultad de Informática	Ingeniería en Computación (en conjunto con la Fac. de	-44	-70
20	28	Masculino	Argentina	Riande Alaco	CENTRO EDUC	Facultad de Cs. Astronómicas y Geofísicas	Ingeniería	-26	-60

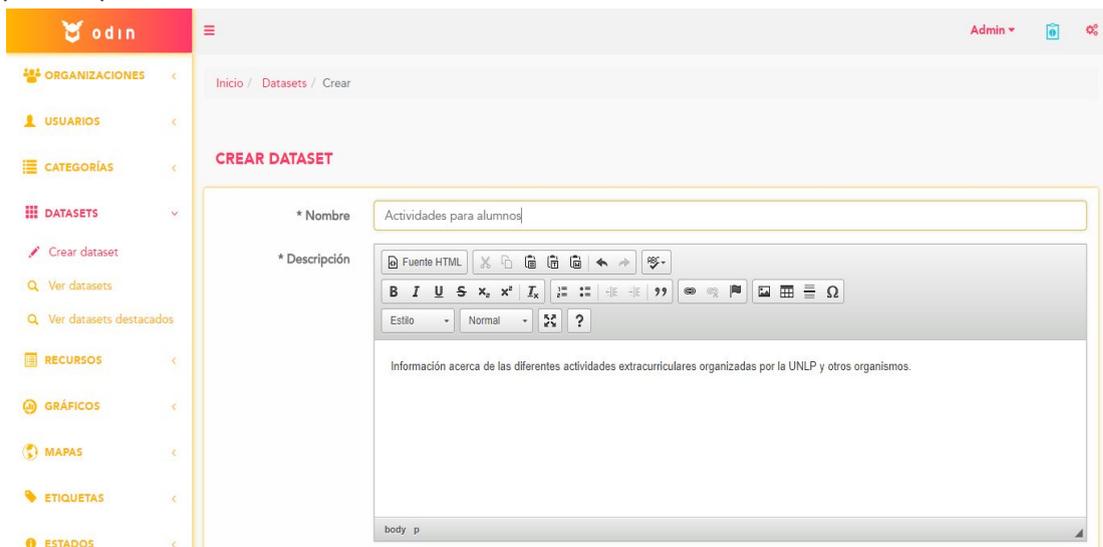
Muestreo de datos de los estudiantes inscriptos de la Universidad Nacional de La Plata.

En consecuencia, se han definido los conjuntos de datos que engloban los recursos mencionados y las visualizaciones que se confeccionarán a partir de los mismos con el fin de disponibilizar de forma concreta las estimaciones y características que dichos datos comprenden.

A partir de esta selección de la información se realizó la confección de las categorías que contendrán los conjuntos de datos definidos: Académico, Espacios, Estudiantes e Institucional. De esta manera, el usuario final del portal de datos abiertos podrá contar con un

pantallazo general de la información presente en la plataforma.

- **Aplicar una licencia abierta:** como se mencionó en el primer capítulo de este documento, los datos deben disponibilizarse bajo condiciones legales abiertas, por tal motivo se decidió optar por la licencia Creative Commons Attribution 4.0 International que posibilita no sólo copiar y redistribuir el material en cualquier medio o formato, sino también adaptarlo y transformarlo para cualquier propósito, incluso comercial.
- **Abrir los datos:** una vez seleccionados los datos a disponibilizar y la licencia con la que se hará, se procedió a la gestión de la información dentro de la herramienta provista por el prototipo desarrollado para realizar la apertura de datos propiamente dicha. Para ello, se crearon las entidades (categorías y organización) asociadas a los conjuntos de datos y los recursos, y se cargaron los valores correspondientes a los metadatos del conjunto de datos para su publicación.



Pantalla de creación de un conjunto de datos del prototipo.

El dataset o conjunto de datos tiene como atributos un nombre, una descripción, las categorías y etiquetas asociadas para posteriores filtros, el usuario responsable de su gestión y datos opcionales de carga para generar nuevos metadatos adicionales que se mostrarán en el catálogo.

A su vez, cuenta con un listado de los conjuntos de datos existentes en la plataforma para facilitar la gestión al usuario administrador y permitirle una fácil manipulación del estado al que corresponde el conjunto de datos.

DATASETS

Crear DATASET

Nombre Estado Autor Categoría

Buscar por Nombre Seleccione un Estado Seleccione un Autor Seleccione un Categoría BUSCAR LIMPIAR

Nombre	Estado	Autor	Categorías	Acciones
Programa de apoyo y contención a alumnos	Publicado	admin	Académico	
Portal web de la UNLP	Publicado	admin	Institucional	
Nómina de desaparecidos y asesinados de la UNLP	Publicado	admin	Estudiantes Institucional	
Actividades para alumnos	Publicado	admin	Estudiantes	
Alumnos	Publicado	admin	Estudiantes Académico	
Elecciones 2017	Publicado	admin	Estudiantes	

Pantalla de listado de los conjuntos de datos del prototipo.

Una vez creado el conjunto de datos, se crearon los recursos pertenecientes al mismo. Esto se puede lograr a través de la subida directa de un archivo físico o mediante un servicio web.

DETALLES PARA ACTIVIDADES PARA ALUMNOS

VER RECURSOS ASOCIADOS + NUEVO RECURSO DESPUBLICAR

Nombre Actividades para alumnos

Descripción Información acerca de las diferentes actividades extracurriculares organizadas por la UNLP y otros organismos

Categorías Estudiantes

Estado Publicado

Creado por admin

Usuario responsable admin

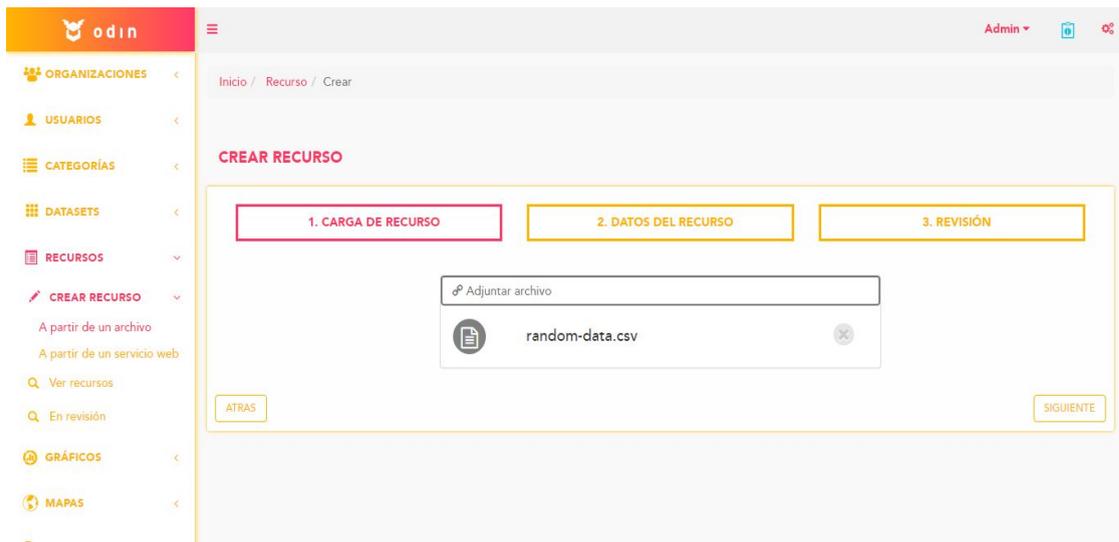
Fecha de creación Jul. 8, 2018 6:51:35 p. m.

Fecha de publicación Jul. 8, 2018 6:51:35 p. m.

Fecha de actualización Jul. 8, 2018 6:51:35 p. m.

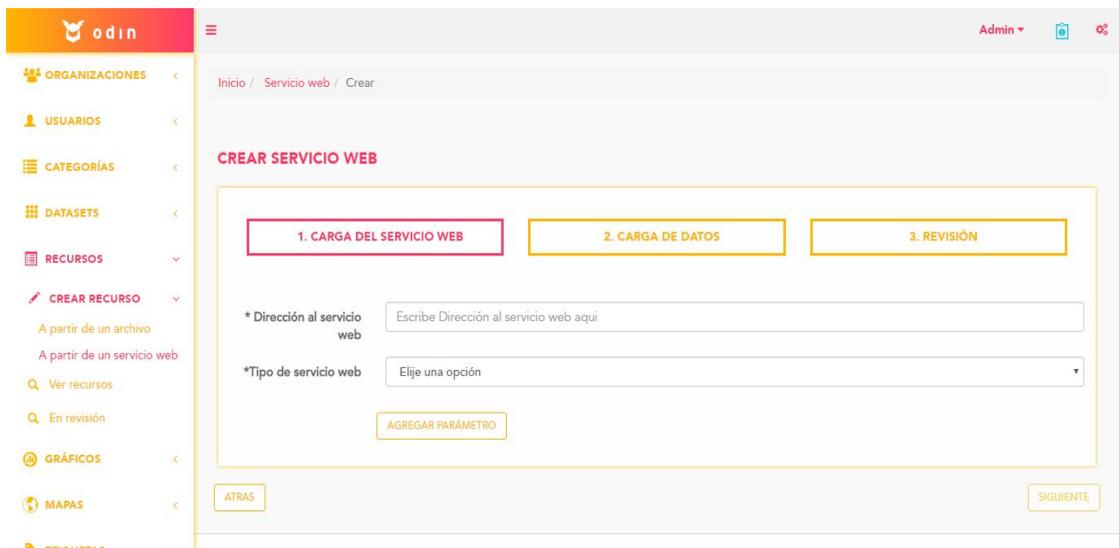
Destacado

Pantalla del detalle del conjunto de datos del prototipo.



Pantalla de creación de un recurso a partir de un archivo físico.

Es importante definir previamente los tipos de archivos permitidos en el sistema, ya que se realiza una comprobación al momento de seleccionar un archivo desde la computadora correspondiente, notificando al usuario en caso de que haya algún inconveniente con el tipo de archivo a subir. Por consiguiente, se creó el tipo de archivo csv para la posterior creación de recursos.

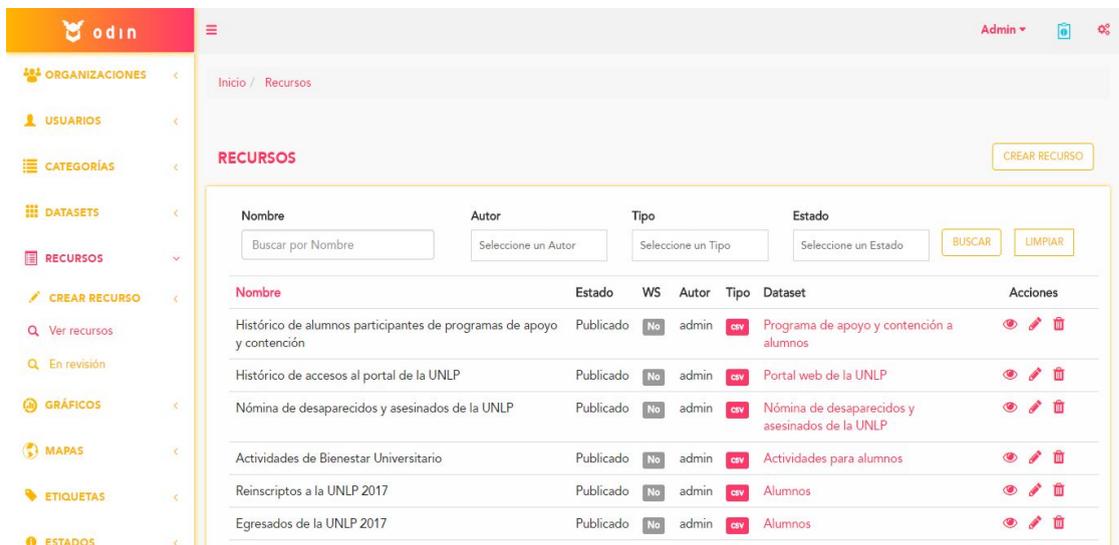


Pantalla de creación de un recurso a través de un servicio web.

En la modalidad de creación de un recurso a través de un servicio web, se cargaron los datos correspondientes a los atributos generales de los servicios web ya sean de tipo REST como de tipo SOAP. Entre ellos: ruta de los datos, ruta a los títulos de las tablas y forma de autenticación en caso de ser REST; y método, nombre de espacio y atributos como cabeceras en caso de ser SOAP.

En ambos casos, se debieron cargar los valores correspondientes a los metadatos de los recursos como el nombre, la descripción, el conjunto de datos al que pertenece, la frecuencia de actualización, la fecha de relevamiento de la información y la organización facilitadora de

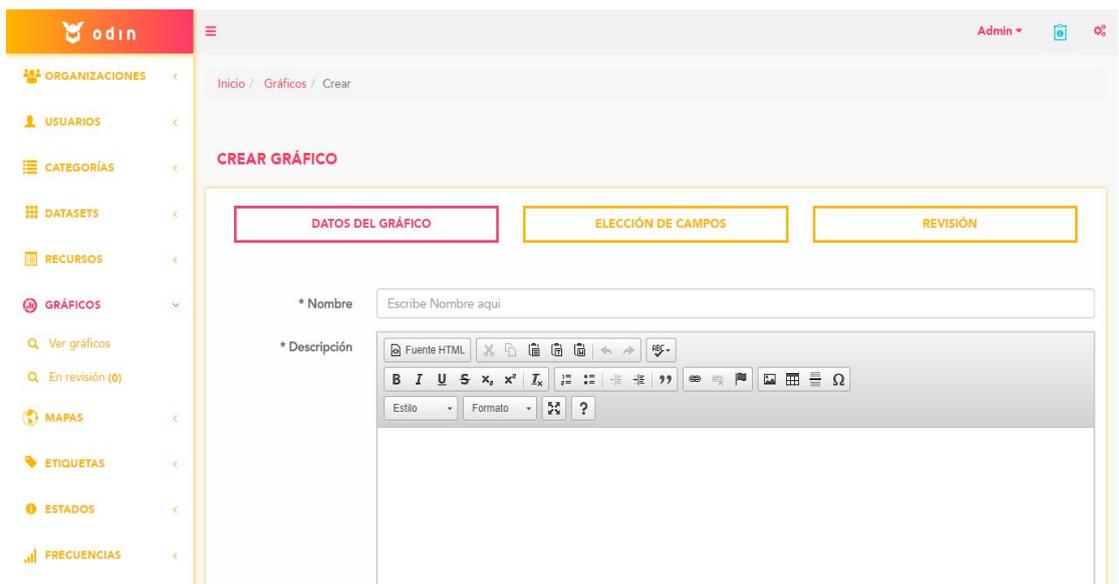
los datos, entre los más relevantes.



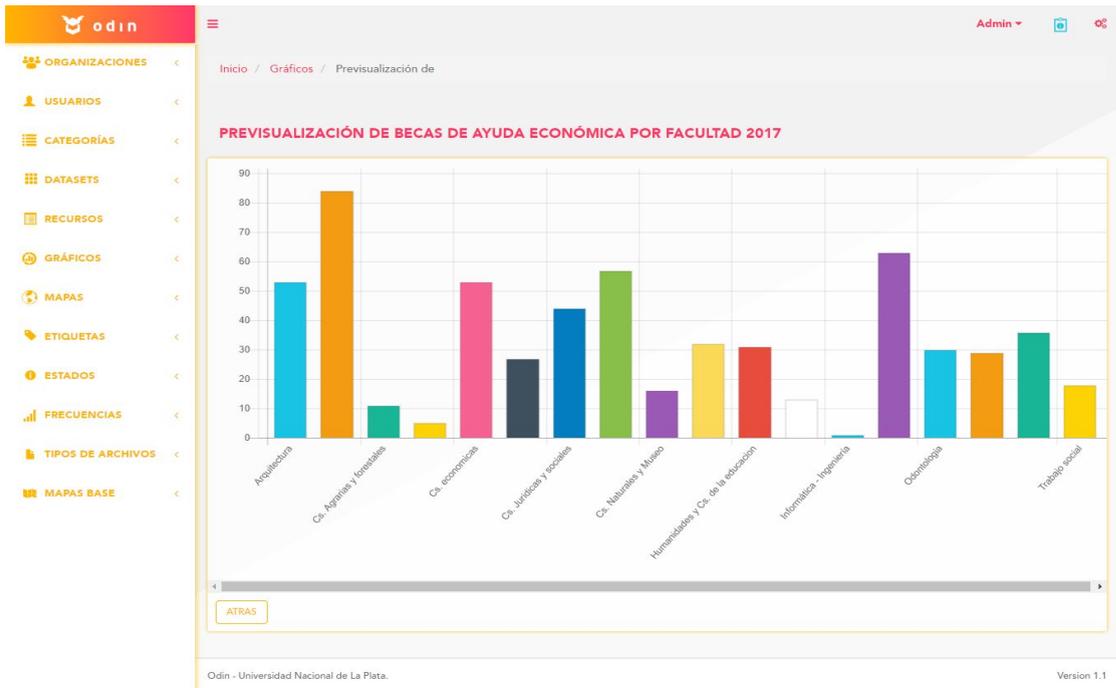
Pantalla de listado de los recursos del prototipo.

A su vez, se cuenta con un listado que facilita la gestión de los recursos existentes en la plataforma y permite acceder de forma sencilla a la publicación y la generación de visualizaciones.

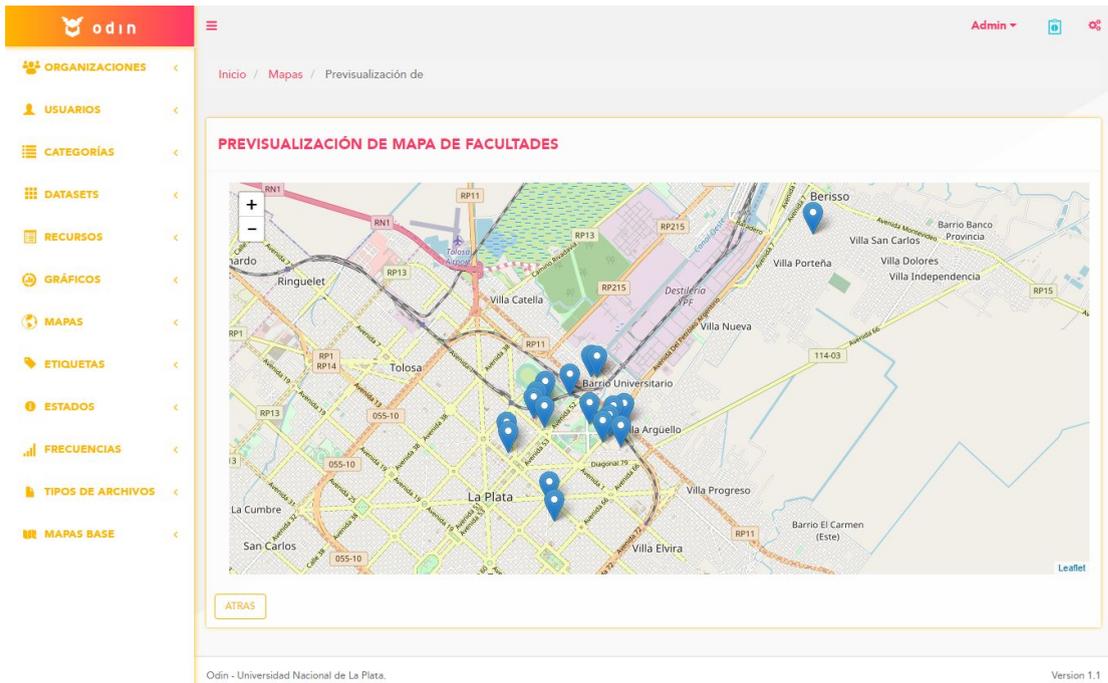
El valor agregado de cualquier plataforma de publicación de datos es la gestión de visualizaciones a partir de los datos presentes en la plataforma, ya que muestra de forma amigable una representación de la información. Por este motivo, el prototipo cuenta con un generador de gráficos y mapas para recursos de tipo csv, xls y xlsx, cargados previamente en el sistema. Para ello, se crearon gráficos y mapas con sus metadatos correspondientes, desde un recurso específico.



Pantalla de creación de un gráfico en el prototipo.



Pantalla de previsualización de un gráfico creado en el prototipo.



Pantalla de previsualización de un mapa creado en el prototipo.

Además, tanto los gráficos como los mapas soportan la vinculación de un enlace embebido para que el usuario pueda utilizar otra herramienta dedicada a este tipo de funcionalidades.

Finalmente, los conjuntos de datos y los recursos creados junto con sus visualizaciones se encuentran publicados en el catálogo de datos.

- **Promover el uso:** una vez realizado el análisis, identificación, selección y gestión de los datos en la plataforma, se debe promover y fomentar la utilización de la información

compartida. Para esto se deben generar vínculos con posibles interesados en la información mediante diversos canales de comunicación como redes sociales, sitios web vinculados a la universidad, secciones exclusivas dentro del propio catálogo de datos; y eventos como mesas de diálogo, hackatones, datatones o ideatones. Es una buena práctica estructurar metas y objetivos a alcanzar con la iniciativa de datos abiertos adoptada. También hay que tomar una postura abierta en cuestiones de permitir la colaboración continua por parte de las comunidades de datos abiertos consolidadas en otros ámbitos con el fin de optimizar los mecanismos de comunicación y difusión.

Informes de usuarios

Luego de haber desarrollado el prototipo de esta Tesina, Odin, se llevaron a cabo pruebas de uso con tres personas de diferentes perfiles para determinar el cumplimiento satisfactorio de las tareas principales, el grado de aceptación de la herramienta y encontrar puntos de mejora. Con tal motivo, se disponibilizó un entorno de prueba sobre el cual se diagramaron tareas específicas de gestión y administración de contenido; uso e interacción con el catálogo de datos; y uso e interacción con la API.

Las pruebas de uso son las siguientes:

Usuario	Sofía, 27 años.	
Perfil	Experta en datos abiertos. Gestión y administración de la información.	
Profesión	Científica de datos.	
Descripción	Se realizaron las tareas el día 12 de Agosto de 2018 de forma virtual mediante el entorno de prueba disponibilizado.	
Herramientas utilizadas	CKAN y Junar.	
Tareas a realizar	Usabilidad. Concepto del conjunto de datos.	Tarea 1: Crear un conjunto de datos.
	Usabilidad. Concepto del conjunto de datos.	Tarea 2: Crear un recurso y asociarlo a su conjunto de datos.
	Usabilidad. Calidad de los datos.	Tarea 3: Crear una visualización del recurso previamente creado.
	Usabilidad. Presentar la información.	Tarea 4: Previsualizar las visualizaciones creadas anteriormente.
	Usabilidad.	Tarea 5: Gestionar alguna entidad asociativa.
	Usabilidad.	Tarea 6: Modificar las configuraciones por defecto de la plataforma.

<p>Resultados obtenidos</p>	<p>Tarea 1: El usuario creó un conjunto de datos correctamente.</p> <p>Tarea 2: El usuario creó y asoció correctamente un recurso.</p> <p>Tarea 3: El usuario creó un gráfico y un mapa asociado al recurso creado previamente de manera exitosa.</p> <p>Tarea 4: El usuario previsualizó el gráfico y el mapa creado previamente de manera exitosa.</p> <p>Tarea 5: El usuario creó correctamente una categoría, una etiqueta, una organización y un tipo de archivo.</p> <p>Tarea 6: El usuario cambió la organización por defecto de la carga de conjunto de datos.</p>
<p>Observaciones</p>	<p>Tarea 1: El usuario encuentra intuitiva la interfaz para la creación del conjunto de datos.</p> <p>Tarea 2: El usuario propone que se permita la creación del recurso en el mismo proceso de creación del conjunto de datos para agilizar la carga masiva de la información. Además, que permita la creación de las entidades asociativas desde un popup u otra alternativa.</p> <p>Tarea 3: Al usuario le pareció muy interesante el módulo planteado. Como mejora a futuro propone tener un gestor de gráficos del estilo que presenta Junar en su herramienta.</p> <p>Tarea 4: Al usuario le resultó simple y concreta la previsualización del gráfico y mapa creado. Además, la previsualización de un recurso de tipo PDF le pareció muy importante, ya que si bien no es recomendable cargar información con dicho formato, por el momento es lo que más se utiliza.</p> <p>Tarea 5: Al usuario le pareció muy interesante como valor agregado frente a las herramientas utilizadas debido a la facilidad de carga y vinculación de las entidades.</p> <p>Tarea 6: El usuario propone la aclaración de la finalidad de cada configuración para que sea más entendible. Valor agregado sobre la cantidad de configuraciones establecidas.</p>

<p>Usuario</p>	<p>Camila, 35 años.</p>	
<p>Perfil</p>	<p>Inexperta en datos abiertos. Consumidora de la información.</p>	
<p>Profesión</p>	<p>Periodista.</p>	
<p>Descripción</p>	<p>Se realizaron las tareas el día 14 de Agosto de 2018 de forma virtual mediante el entorno de prueba disponibilizado.</p>	
<p>Herramientas utilizadas</p>	<p>-</p>	
<p>Tareas a realizar</p>	<p>Usabilidad.</p>	<p>Tarea 1: Ingresar a la información de un</p>

	Concepto del conjunto de datos. Presentar la información.	conjunto de datos.
	Usabilidad.	Tarea 2: Descargar un conjunto de datos.
	Usabilidad.	Tarea 3: Descargar un recurso.
	Usabilidad. Presentar la información.	Tarea 4: Previsualizar un recurso.
	Usabilidad.	Tarea 5: Buscar un conjunto de datos.
	Usabilidad.	Tarea 6: Filtrar el listado de conjuntos de datos.
	Aspecto legal y licencia.	Tarea 7: Informarse sobre las políticas de uso de los datos.
	Usabilidad.	Tarea 8: Hacer una consulta de contacto.
Resultados obtenidos	<p>Tarea 1: El usuario encontró la página con la información del conjunto de datos correctamente.</p> <p>Tarea 2: El usuario descargó un conjunto de datos correctamente.</p> <p>Tarea 3: El usuario descargó un recurso correctamente.</p> <p>Tarea 4: El usuario previsualizó la tabla, el gráfico y el mapa asociados a un recurso correctamente.</p> <p>Tarea 5: El usuario realizó una búsqueda sobre el contenido del catálogo de datos correctamente.</p> <p>Tarea 6: El usuario realizó un filtro por etiquetas, formatos y organizaciones correctamente.</p> <p>Tarea 7: El usuario encontró la página con la información de las políticas de uso de los datos correctamente.</p> <p>Tarea 8: El usuario realizó una consulta a través del formulario de contacto correctamente.</p>	
Observaciones	<p>Tarea 1: El usuario propone distinguir la página del listado de conjuntos de datos con algún título para diferenciarla del listado de los recursos que tiene dicho conjunto. Además, le resulta de gran valor la inclusión de la guía de datos en el panel principal del conjunto de datos.</p> <p>Tarea 2: Al usuario le resulta interesante el archivo .txt con la información de todo el conjunto de datos que se incluye en el archivo .zip generado para la descarga.</p> <p>Tarea 3: Al usuario le parece de gran valor la posibilidad de descargar un recurso csv, en xls o xlsx, o alguna de sus combinaciones.</p> <p>Tarea 4: Al usuario le resulta interesante lo planteado en los gráficos y</p>	

	<p>mapas asociados a los recursos. Le permite tener noción de la información que puede descargar.</p> <p>Tarea 5: Para el usuario, un buscador general en una plataforma le resulta poco interesante ya que no sabe cuál es el algoritmo que realiza dicha búsqueda. De todas formas, le pareció correcta la búsqueda sobre el conjunto de datos. Propone que la búsqueda impacte, también, en la información de los recursos de la plataforma.</p> <p>Tarea 6: Al usuario le resulta amigable la búsqueda autocomplete de los valores de las etiquetas, formatos y organizaciones para realizar los filtros sobre el listado de conjuntos.</p> <p>Tarea 7: Al usuario le resulta completa y concisa la información encontrada.</p> <p>Tarea 8: Sin comentarios.</p>
--	--

Usuario	Javier, 26 años.	
Perfil	Desarrollador.	
Profesión	Desarrollador Fullstack.	
Descripción	Se realizaron las tareas el día 15 de Agosto de 2018 de forma virtual mediante el entorno de prueba disponibilizado.	
Herramientas utilizadas	Angular, Node.js, Postman	
Tareas a realizar	Usabilidad.	Tarea 1: Usar endpoint para listar los recursos de la plataforma.
	Presentación de la información.	Tarea 2: Usar endpoint para ver la información de un recurso.
	Autodescriptibilidad.	Tarea 3: Navegar la API mediante la metadata recibida.
	Usabilidad.	Tarea 4: Descargar un recurso mediante el atributo 'url' de su modelo desde la API.
	Usabilidad.	Tarea 5: Utilizar la API para buscar un conjunto de datos.
	Usabilidad.	Tarea 6: Filtrar el listado de conjuntos de datos mediante la API.
Resultados obtenidos	<p>Tarea 1: El usuario listó los recursos mediante el endpoint encontrado leyendo la documentación técnica de la plataforma.</p> <p>Tarea 2: El usuario encontró un recurso que le pareció interesante y utilizó el endpoint correspondiente para ver la información contenida.</p>	

	<p>Tarea 3: El usuario observó la metadata, y mediante ella pudo navegar la API.</p> <p>Tarea 4: El usuario listó los recursos, eligió uno, copió el link de descarga, lo pegó en una pestaña nueva del navegador y descargó el archivo.</p> <p>Tarea 5: Siguiendo los estándares de la documentación técnica, el usuario pudo buscar un conjunto de datos, tomando como patrón de búsqueda una palabra en la descripción del mismo.</p> <p>Tarea 6: El usuario filtró por el campo 'descripción' los set de datos que contenían la palabra ingresada y encontró el cual buscaba.</p>
Observaciones	<p>Tarea 1: El usuario propone una manera distinta, preferentemente automatizada, para obtener el token necesario para utilizar la API.</p> <p>Tarea 2: Al usuario le llamó la atención la forma de almacenar los ID y propuso utilizar autoincrementales en lugar de un hash.</p> <p>Tarea 3: El usuario destacó la inclusión de la metadata e hizo hincapié en que le resultaría muy útil a la hora de desarrollar una aplicación.</p> <p>Tarea 4: El usuario no sabía que ese recurso estaba disponible en múltiples formatos.</p> <p>Tarea 5: Al usuario le pareció bueno el cumplimiento de estándares entre endpoints de la plataforma.</p> <p>Tarea 6: Al usuario le costó en un principio entender la diferencia entre la búsqueda y el filtro</p>

En conclusión, la herramienta tuvo un grado de aceptación alto. Los usuarios de testeo la encontraron amigable, intuitiva y de uso fácil tanto como usuario administrador, consumidor del catálogo de datos y desarrollador. No se necesitó realizar una explicación extensa de cómo utilizar la herramienta. Por otra parte, si bien en cuanto a usabilidad la herramienta logró la aprobación de los usuarios, estos plantearon mejoras y ajustes con el fin de escalar a nuevos módulos que les permitan no solo agilizar los procesos que cada uno requiere, sino también para fomentar y perfeccionar el uso de una herramienta de código abierto a medida.

Resumen

En conclusión, la iniciativa de datos abiertos pone en manifiesto un esquemático número de pasos y procesos para lograr finalmente la apertura de la información propiamente dicha y acaparar la atención de una comunidad enfocada en el consumo y aprovechamiento de los datos abiertos que va creciendo día a día. No solo se plantea el análisis de herramientas alternativas para la difusión de la información que, sin lugar a dudas es la actividad principal de esta política, sino que también conlleva un análisis sobre la información que la entidad vinculada y los beneficios que la apertura de esta pueden generar a quienes la consuman mediante alguno de los mecanismos de difusión que la plataforma de publicación de datos abiertos disponga. A esto se le debe sumar el aporte significativo

por parte de los consumidores de la información que interactúan con la misma, ya sea por pertenecer al equipo de datos abiertos de la entidad disponibilizadora o generadora de la información, como por pertenecer a una comunidad de desarrolladores, periodistas, alumnos o cualquier ciudadano que muestre interés por el valor de la información que está analizando.

Sin lugar a dudas, la utilización de plataformas de publicación de datos abiertos como medio principal para la comunicación, transparencia y generación de proyectos innovadores es el motor de una iniciativa a nivel mundial que va tomando mayor relevancia con el paso de los años.

Referencias

[1] https://www.argentina.gob.ar/sites/default/files/2._kit_de_datos_abiertos.pdf, fecha de consulta: 28 de Julio 2018.

Propuestas para futuros desarrollos

Con el objetivo de mejorar la herramienta desarrollada para esta tesina se proponen ideas para futuros desarrollos:

- Esquematizar y realizar una federalización de la herramienta para las distintas universidades que permita realizar búsquedas generales de información y que localice la universidad que la disponibiliza. Esto podría agilizar la gestión de la información con características similares de cada una de las universidades y facilitaría al usuario consumidor la navegación sobre los datos disponibles en el catálogo de datos.
- Estandarizar la normalización de los datos a través de una entidad centralizadora o en común acuerdo entre cada una de las universidades con el fin de agilizar y perfeccionar la interacción, integración y combinación de los datos disponibilizados por las mismas.

Anexo

A continuación se listan los links más importantes del prototipo desarrollado:

- Repositorio del prototipo: <https://github.com/odin-unlp>
- Documentación técnica: <https://odin-unlp.github.io/odin-tech-docs/>
- Url del ambiente de testeo del catálogo de datos: <http://206.189.190.2/>
- Url del ambiente de testeo del administrador: <http://206.189.190.2/admin>
- Credenciales de acceso del administrador: usuario: admin - contraseña: 123
- Url del ambiente de testeo de la API: <http://206.189.190.2/api>

Conclusiones

A lo largo de los últimos años, las iniciativas de datos abiertos han ido tomando fuerza en varios sectores a nivel global. El ámbito de mayor impacto es el que se conoce como gobierno abierto y es del que se desprenden varios de los conceptos, mencionados a lo largo de todo el documento, que son tomados por áreas científicas y universitarias, entre otras, como base estructural para la promoción de sus políticas de datos abiertos. Casos como el mapa humanitario creado a partir de la cruza de datos abiertos entre entidades gubernamentales y no gubernamentales para distinguir las zonas afectadas por el Ébola en Sierra Leona y mejorar los servicios sanitarios entorno al problema; la creación de un sitio para gestionar las diferentes ayudas y medios de apoyo, por un grupo de personas pertenecientes a la organización Codeando México para hacer frente al sismo ocurrido en dicho país en el año 2017; y la creación de un sitio de búsqueda inmobiliaria de Properati, que utiliza los datos abiertos para generar mapas de calor o establecer puntos de conveniencia en la compra o alquiler de un inmueble y, además, pone a disposición los datos generados dentro de su plataforma, ponen en manifiesto los grandes beneficios y el impacto de llevar a cabo una iniciativa de datos abiertos independientemente del área en que se encuentre.

Si bien las iniciativas de datos abiertos en el ámbito universitario tienen sus comienzos en estos últimos años, varios son los casos de avance en la temática como son la Universidad de Oxford (UK), Universidad de Southampton (UK), Universidad de Alicante (España), Universidad de Pablo de Olavide (España), Universidad de Waterloo (Bélgica) y Universidad de Münster (Alemania), entre las más destacadas debido al nivel de transparencia logrado en sus instituciones y los servicios brindados a sus estudiantes y docentes, tanto en lo relativo a la enseñanza como al aspecto cotidiano de interacción entre los mismos y sus áreas académicas. Es importante destacar el desprendimiento de conceptos como investigación abierta y acceso abierto que se desprende de lo que se conoce como universidad abierta, ya que muchos de los portales de compartición de datos de investigación surgen de las universidades con el fin de agilizar las tareas de los docentes a cargo de los programas de educación y a los investigadores y estudiantes que participen en los diferentes proyectos incentivados por la propia universidad.

Se puede ver, a lo largo de todo el documento, la brecha existente entre la mayoría de países norteamericanos y europeos respecto de los países latinoamericanos en relación a los niveles de transparencia y escalas de apertura de datos en sus instituciones gubernamentales. No obstante, los países latinoamericanos se encuentran en un compromiso de desafíos en el tema que pretenden mejorar estos niveles y escalas en los próximos años y mantener su crecimiento en forma paulatina y constante. Sin embargo, en lo que se conoce como universidad abierta se encuentran muy lejos de lo pretendido, ya que son pocos los países que comenzaron con iniciativas de dicha índole, como la Universidad Nacional Autónoma de México y la Universidad Central del Ecuador que ponen a su disposición portales de datos abiertos en formatos abiertos. Universidades de Brasil, Chile y Argentina, por otra parte, muestran en sus sitios web institucionales información sobre la transparencia y documentos en formatos no abiertos con información vinculada a la administración en sus instituciones.

La importancia de llevar adelante la iniciativa de datos abiertos en la Universidad Nacional de La Plata (UNLP) determinó la generación de un proceso de apertura de datos y la creación de un prototipo que permita su gestión y publicación. Dicho proceso contempla la investigación y selección de los datos administrados por la UNLP, la generación de datos aleatorios que permitan resaltar los beneficios posibles de la utilización de un portal de datos abiertos y de la implementación de una iniciativa de esta índole, el análisis y selección de una licencia abierta que abastezca las necesidades

de los consumidores y productores de la información, la apertura de datos propiamente dicha a través del prototipo desarrollado para esta Tesina y la difusión de los datos. Es importante destacar el previo análisis de las herramientas que se encuentran en el mercado para la disponibilización de los datos y de las herramientas tecnológicas que fueron utilizadas para la creación del prototipo final.

La herramienta de publicación de datos abiertos contempla la gestión y categorización de la información, y un amplio repertorio de puntos de interacción con el consumidor final de los datos, ya sea para la descarga o visualización de la misma, como los canales de comunicación. A lo largo de la implementación y el proceso de análisis de las herramientas y gestión de la apertura se pusieron en práctica materias vistas a lo largo de la carrera licenciatura en sistema, tales como: Proyecto de Software, Bases de datos, Ingeniería de Software, Aspectos Legales, Tecnologías Aplicadas para Business Intelligence, Redes y Comunicaciones, Cloud Computing, Desarrollo Seguro de Aplicaciones y Diseño de Interacciones en Aplicaciones Móviles.

Por último, se ha logrado la aceptación de la herramienta por parte de los productores de la información, los usuarios del catálogo de datos y los perfiles desarrolladores debido a que dicha herramienta cuenta con una API Restful para su consumo programático. Más allá del visto bueno de estos usuarios y la generación de los informes finales, se han recabado puntos de mejora de la plataforma con el fin de lograr un crecimiento futuro que abastezca las necesidades de los usuarios. Es importante mencionar que este resulta el comienzo de una iniciativa de datos abiertos en la Universidad Nacional de La Plata que se espera tome fuerza con el paso del tiempo.

Bibliografía

<https://hipertextual.com/archivo/2014/05/diferencias-software-libre-y-open-source/>

<https://www.europeandataportal.eu/en/using-data/benefits-of-open-data>

<https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information>

<https://www.opendatasoft.es/2017/01/04/todo-sobres-datos-abiertos/>

<https://www.bbva.com/publicaciones/datos-abiertos-big-data-al-servicio-de-los-ciudadanos/>

<https://www.theguardian.com/public-leaders-network/2014/apr/15/big-data-open-data-transform-government>

<http://datos.gob.es/es/noticia/datos-abiertos-big-data-y-gobierno-abierto-diferentes-tipos-de-datos>

<https://opendatacommons.org/guide/>

http://guidance.data.gov.uk/five_stars_of_openness.html

<https://idatosabiertos.org/datos-para-la-rendicion-de-cuentas-una-gran-brecha/>

<https://www.viavansi.com/blog-xnoccio/es/datos-abiertos-en-latinoamerica/>

<https://distintaslatitudes.net/lo-mejor-y-lo-peor-de-america-latina-en-datos-abiertos>

<https://idatosabiertos.org/ecosistemas-de-datos-abiertos-en-america-latina-el-aporte-de-ilda/>

<https://idatosabiertos.org/el-surgimiento-de-america-latina-abierta/>

<https://idatosabiertos.org/the-emerging-impact-of-open-data-in-latin-america/>

<http://odimpact.org/case-brazils-open-budget-transparency-portal.html>

<http://odimpact.org/case-united-kingdoms-transport-for-london.html>

<http://odimpact.org/case-paraguays-dengue-prediction.html>

<http://odimpact.org/case-mexicos-mejora-tu-escuela.html>

<http://odimpact.org/case-uruguays-a-tu-servicio.html>

<http://www.excelsior.com.mx/nacional/2017/07/17/1176162>

<https://developer.ibm.com/dwblog/2017/what-is-big-data-insight/>

<http://www.ibmbigdatahub.com/infographic/four-vs-big-data>