

¿CÓMO DESARROLLAR BIENES Y SERVICIOS PÚBLICOS CON DATOS?

Diego Ferreyra

La relación entre las tecnologías de la información y la noción de estado en términos de institución tiene pasado tan frondoso como el estado mismo. Existen autores (De Santillana, 1965; Wright, 2007) que incluso consideran que las expresiones tan antiguas como las primeras ciudades Estado, sólo pueden explicarse a partir del surgimiento de la escritura y la posibilidad de externalizar en soporte perdurable reglas, atributos o derechos, permitiendo la consolidación de grupos, castas y estructuras jerárquicas estables, necesarias para la consolidación de las ciudades estado como tales.

Así como las primeras ciudades Estado resultaron viables a partir de la escritura, es posible aventurar la idea de que la paulatina incorporación al procesamiento de datos ocurrida en el siglo XIX de procesos automatizados y fuerza motriz, es condición de posibilidad del surgimiento y desarrollo de los grandes conglomerados urbanos durante el siglo XIX.

Aun en la noción moderna de estado-nación es posible hallar una vinculación entre las tecnologías de la información y sus condiciones de posibilidad, siendo que los estados basados en constituciones son justamente formas de organización social basadas en un dispositivo tecnológico específico: un tipo específico de documento denominado constitución.

En los últimos cincuenta años es posible identificar una relación cada vez más estrecha entre la condiciones de institucionalidad de una estruc-

tura social de gestión y sus capacidades para construcción, consolidación y gestión de datos. Definida alguna vez como una nueva “tecnología intelectual”, la construcción de datos, su disponibilidad y su potencial uso es, en la actualidad, una dinámica omnipresente y transparente que atraviesa esferas, fronteras y territorios de todo tipo. Imaginemos la recepción de una cuenta de teléfono mecanografiada o de un resumen de tarjeta de crédito manuscrito: de inmediato una sensación de desconfianza hacia la información allí vertida y documentada se apodera de nosotros... ¿serán ciertos los datos?, ¿puedo confiar en que esos garabatos manuscritos son mis consumos del último mes?, ¿será auténtica la factura?, ¿podré presentarla como comprobante? Para nuestra tranquilidad, logramos comunicarnos con el servicio de atención al cliente detallado en el margen inferior de las facturas y allí evacuan nuestras dudas acerca de la documentación recibida. Pero si acaso no nos brindan un número de reclamo que documente nuestra comunicación, ¿podemos realmente quedarnos tranquilos? Como vemos, es muy difícil imaginar un proceso de gestión institucional desprovisto de mecanismos sistemáticos, formales y estructurados de gestión de datos. Asimismo, en la pequeña anécdota anterior, es posible identificar una clara relación entre las capacidades tecnológicas para la gestión de datos y las condiciones de legitimidad y autoridad del dispositivo de gestión. Justamente, la omnipresencia de este tipo de dinámicas genera un efecto de transparencia que acaba por invisibilizar el conjunto de dispositivos, actores, políticas, metodologías y tecnologías que intervienen. Todo este universo se torna opaco y repentinamente visible cuándo irrumpe la falla: un corte de luz, un boleto de avión hecho a mano en una agencia de turismo o la caída de la una red de celulares nos presentan en un único instante la totalidad de factores que participan de cada uno los procesos de gestión basados en datos.

Aquellos servicios, prestaciones y capacidades cuya presencia, fiabilidad, utilidad y permanencia damos por cierta, generan las condiciones

de posibilidad para el desarrollo de nuevos servicios y prestaciones: los caminos definen los tipos de medios de transporte como los electrodomésticos presuponen la electricidad. A este tipo de relación, de capacidad y transparencia la denominamos infraestructura. Como vimos, la infraestructura no sólo es un agente de legitimidad y autoridad, es un moderador de formas de participación e intervención social: cada infraestructura propone una arquitectura específica de participación, distribuye roles y posibilidades. Ahora bien, surge entonces una pregunta: ¿qué significan, en el contexto del Estado y en términos de política pública, los datos como infraestructura? O de manera más puntual: ¿cómo desarrollar bienes y servicios públicos con datos?

Los datos

En este punto, una vez resueltos a abordar la problemática del desarrollo de bienes y servicios públicos basados en datos, resulta relevante formularnos una pregunta que no por elemental resulta menos significativa: ¿de dónde vienen los datos?, ¿pueden ser recogidos en la naturaleza?, ¿se cultivan?, ¿crecen en todos los climas?

Los datos, en nuestro contexto, son un instrumento capaz de delimitar y describir una entidad o fenómeno según un determinado marco teórico y utilizando una metodología definida y formal. Permiten segmentar una realidad continua en elementos discretos que resultan significativos en relación a marcos teóricos, metodológicos e instrumentales específicos. Los datos son el insumo primario de cualquier proceso de análisis, entendiendo el análisis justamente como un proceso que permite separar y definir entidades y establecer sus cualidades diferenciales con respecto a otras entidades o con respecto a la misma entidad a través del tiempo.

Una vez elaborados los datos, se nos presenta una nueva particularidad propia de ellos: no alcanza con disponer de datos para simplemente poder utilizarlos, es necesario estructurarlos y codificarlos.

Una vez elaborados los datos, resulta necesario seguir ciertos lineamientos y estrategias para que los mismos puedan ser utilizados. Es decir, más allá de disponer de datos, para poder consolidar una infraestructura de servicios basada en datos los mismos deben estar provistos de condiciones de portabilidad con el fin multiplicar realmente los escenarios de uso.

En tal sentido, con el fin de adoptar una concepción amplia y comprehensiva de la noción de portabilidad, resulta útil la formalización conceptual realizada en el marco del Open Language Archives Community (OLAC). La misma considera que se encuentran cumplimentadas las condiciones de portabilidad cuando la información es codificada de manera tal que su uso, tratamiento y gestión pueda ser realizado:

- 1.** utilizando diferentes aplicaciones informáticas;
- 2.** utilizando diferentes infraestructuras informáticas;
- 3.** según diferentes comunidades de prácticas;
- 4.** en función de propósitos diversos;
- 5.** a través del tiempo.

De esta manera, la propuesta de portabilidad del denominado marco OLAC está centrada en promover esquemas de codificación de datos que permitan usos no previstos, por parte de usuarios no previstos y en contextos no previstos en un primer momento por quienes han elaborado los datos. Justamente, uno de los atributos más valiosos de las obras de infraestructura como tal es su capacidad para tolerar y viabilizar un amplio abanico de estrategias de uso y explotación, a veces incluso no previstas en el esquema inicial del servicio.

Al igual que las autopistas o los trenes, para alcanzar un modelo de explotación parangonable a la noción de infraestructura no alcanza solamente con ofrecer amplias capacidades o prestaciones críticas, los datos deben ser provistos de manera persistente, uniforme y homogénea, es decir, deben ofrecerse en condiciones de previsibilidad constante para que otros actores sociales puedan diseñar y articular servicios basados en la infraestructura, en esta caso, provista por el Estado.

Hasta aquí hemos recorrido el camino de los datos, hemos visto que son elaboraciones dependientes de modelos teóricos y metodológicos específicos, que requieren de determinadas condiciones para facilitar sus condiciones de reutilización y explotación. En el siguiente apartado analizaremos esas condiciones en general, y en particular aquellas que resultan diferenciales en ambientes digitales.

Los datos en ambientes digitales

Tal como vimos en el anterior apartado, para poder desarrollar bienes y servicios públicos sobre la base de datos, es necesario proveer a los mismos de condiciones portabilidad, persistencia y uniformidad de acceso y disponibilidad que faciliten estrategias de reutilización de los mismos. Para lograr estas condiciones básicamente resulta crítico establecer tres mecanismos articulares que habrán de cargar con gran parte de la responsabilidad de proveer las condiciones antes mencionadas. Estos mecanismos son pues:

- mecanismos para identificar datos: métodos específicos para denominar e identificar de manera unívoca cada dato o conjunto de datos;
- mecanismos para localizar datos: métodos que permitan localizar y acceder a un dato en base a su anterior identificación;
- mecanismos para representar datos: esquemas y herramientas

que permitan, una vez identificados y localizados los datos, codificar y representar a los mismos para su uso y comunicación.

La implementación de estos mecanismos en ambientes digitales conlleva una peculiaridad: en ambientes digitales los datos para ser legibles para humanos primero deben ser legibles para una máquina y sus procesos algorítmicos. Es decir, en ambientes digitales, con miras a la conformación de una infraestructura de servicios basados en datos, resulta crucial implementar estrategias tecnológicas orientadas a maximizar sus condiciones de compatibilidad con otras tecnologías. Quizás resulte paradójico, pero de esta manera se maximizan las posibilidades de comunicación y difusión social. Un servicio de datos susceptible de ser utilizado en computadoras, televisores y celulares resultará más accesible y probablemente dispondrá de un mayor alcance y difusión que un servicio de datos que sólo puede ser utilizado a través de una computadora o un programa informático específico.

Si bien existe un variado espectro de estrategias en uso en ambientes digitales para lograr este conjunto de requisitos (identificación, localización, representación), en el contexto de la WWW una de las estrategias consideradas como un modelo de buenas prácticas por diversas organizaciones y asociaciones profesionales (Consortio W3, Open Knowledge Foundation, UNESCO, etc), se basa en la utilización de tres componentes técnicos que ya disponen de un uso masivo y extendido ya que conforman parte de la arquitectura de base de la WWW: nos referimos a los llamados identificadores uniformes de recursos (URI, siglas de Uniform Resource Identifier), los localizadores uniformes de recursos (URL, siglas de Uniform Resource Locator) y al lenguaje de marcas XML (siglas de eXtended Markup Language). A continuación se describirán estas tecnologías y podremos ver hasta qué punto ya están en uso en muchas situaciones cotidianas.

- URI: Uniform Resource Identifier

La traducción de la sigla URI es identificador uniforme de recurso, es decir, es una sintaxis que define cómo redactar identificadores de recursos en el contexto de Internet. Por ejemplo, los números telefónicos también tienen una sintaxis, hay que anteponer números específicos para llamar a un celular o hay números especiales que representan y funcionan como códigos de áreas, etc. De la misma manera, existe una sintaxis específica que permite redactar código de identificación de recursos que nos permiten identificar de manera unívoca cualquier recurso en internet. De esta manera, cada URI se conforma de una identificación del esquema en el que opera el identificador (por ejemplo: http, URN, mailto), luego, precedido de una barra doble (/), se consigna el proveedor primario de autoridad para ese identificador (ejemplo <http://gobiernolocal.gov.ar>), y luego, separados por una barra simple (/), las posibles subdivisiones jerárquicas del identificador (<http://gobiernolocal.gov.ar/atlas/>). Por ejemplo, todas las direcciones de correo electrónico son URIs, y también las direcciones de los sitios web, las URLs, son un tipo de URI.

- URL: Uniform Resource Locator

La traducción de la sigla URL es localizador uniforme de recursos. Básicamente es un tipo de URI (por lo tanto también dispone de un esquema, autoridad, ruta jerárquica, etc), pero además, a través del esquema (http) se encuentra asociado a un protocolo específico encargado de instrumentar la localización y recuperación del recurso. De esta manera, cuando utilizamos una URL para acceder a un sitio, estamos utilizando una URI, pero además, estamos recurriendo a un protocolo específico para que una vez identificado el recurso en el contexto de Internet, lo recupere y nos permita acceder a él.

- XML: eXtended Markup Language

La traducción de la sigla XML es lenguaje de marcas extensible. Un

lenguaje de marcas es básicamente un lenguaje que utiliza un conjunto delimitado y específico de caracteres para definir y estructurar datos en texto plano. Un ejemplo cotidiano y corriente podría ser los signos de puntuación con respecto al lenguaje escrito. Por ejemplo, en el presente documento, damos por cierto un punto luego de una palabra, delimita y define el final de una oración. En base a una norma anterior, denominada SGML (Standard Generalized Markup Language), el XML utiliza los caracteres < y > para definir los límites de cada elemento (una oración, un párrafo, etc). En el contexto de este tipo de lenguajes, cada delimitador se denomina etiqueta. A diferencia del punto en las oraciones, las etiquetas pueden definir y estructurar todo tipo de elementos y datos. Por ejemplo, en el contexto de una archivo XML, podríamos codificar una oración de la siguiente manera: <oracion>Esta es una oración</oracion> o un dato de población de la siguiente manera: <poblacion>34000</poblacion>. Cada sentencia delimitada por etiquetas conforma un nodo. Para expresar jerarquía o inclusión, es posible anidar un nodo en otro, por ejemplo, el nodo <oracion>, seguramente estará codificado dentro del nodo <parrafo>. De esta manera, si acaso fuera necesario expresar la población total de Roque Pérez según el INDEC, esta podría ser la estructura resultante en XML:

```
<gobiernolocal>
  <nombre>Roque Perez</nombre>
  <CodigoIndec>60693</CodigoIndec>
  <indec>
    <censo1991>
      <poblacion>9639</poblacion>
    </censo1991>
    <censo2001>
      <poblacion>10902</poblacion>
    </censo2001>
    <censo2010>
      <poblacion>12513</poblacion>
    </censo2010>
  </indec>
</gobiernolocal>
```

Tal como es posible constatar, a través de esta sintaxis es posible definir y codificar todo tipo de estructuras de datos. La definición de las etiquetas que componen una estructura de datos se denominan vocabularios, en el sentido en el cuál un vocabulario delimita y define el tipo de “cosas” que pueden existir, los vocabularios en el contexto de un archivo XML definen la entidades y su forma de estructuración. Dichos vocabularios explicitan y formalizan la manera en que están estructurados cada conjunto de datos y facilitan la gestión clara y diferenciada de estructura y datos.

Este conjunto de factores, como ser la utilización de texto plano o la posibilidad de definir y codificar todo tipo de datos, permite maximizar las condiciones de portabilidad tecnológica de cualquier dato que sea expresado y representado a través de un archivo XML.

Hasta aquí se ha descrito de manera amplia una estrategia viable para el desarrollo de servicios basados en datos promoviendo condiciones de portabilidad e interoperabilidad óptimas. En el siguiente apartado se abordará la aplicación de lo visto hasta el momento en el ámbito de gobierno.

Datos para gobierno abierto

En el ámbito de gobierno, existe un importante número de iniciativas tendientes a lograr la exposición pública de datos a través de medios digitales. Para la presente síntesis resulta interesante considerar una propuesta de lineamientos elaborada a partir de un taller de reflexión y elaboración realizado en el año 2007 en Sabatopol (Estados Unidos)¹ en el que participaron especialistas y entusiastas del gobierno abierto. Los lineamientos fueron cristalizados en ocho principios que, en la actualidad, se han convertido en un estándar de facto para el análisis y evaluación de servicios basados en datos en contextos de gobierno.

Ocho principios para datos gubernamentales abiertos

1. **Completos.**

Todos los datos públicos han de estar disponibles. Son públicos todos los datos con excepción de los que tienen limitaciones por cuestiones de seguridad, confidencialidad o privilegios especiales.

2. **Primarios.**

Los datos han de estar disponibles con el máximo nivel de desagregación posible, no agrupada o modificada.

3. **A tiempo.**

Los datos estarán disponibles lo más rápidamente posible tras su obtención para que no pierdan su valor.

4. **Accesibles.**

Los datos estarán disponibles para la gama más amplia de usuarios y el espectro más amplio de propósitos.

1 Open Government Working Group Disponible en: https://public.resource.org/open_government_meeting.html Fecha de consulta 2012-10-03

5. **Tratamiento automatizado.**

Los datos se suministrarán de forma que se permita su tratamiento automatizado. Los datos deben poder ser descargados en forma completa, sin procesamiento, por medio de un mecanismo automatizado.

6. **No discriminación.**

Los datos estarán disponibles para cualquier persona, sin requisito previo de registro.

7. **Formato no propietario.**

Los datos estarán disponibles en formatos no propietarios.

8. **Sin licencia.**

Los datos no estarán sometidos a ninguna regulación de patente, marca registrada o secreto comercial.

El portal gobiernolocal.gov.ar como proveedor de datos abiertos

El presente artículo comenzaba con la pregunta acerca de cómo hacer bienes y servicios públicos con datos, luego de establecer su origen y propuestas de buenas prácticas para su identificación, representación, acceso y disponibilidad. En este punto ha llegado el momento de presentar la dinámica a través de la cual el portal gobiernolocal.gov.ar expone los datos que gerencia acerca de los gobiernos locales en la Argentina.

En particular, el portal ha optado por exponer los datos siguiendo los principios antes expuestos a través de un esquema específico de servicios web. Un servicio web es una vía para exponer datos en un formato específico y según un protocolo o una sintaxis específica de consulta. Tal como lo reseña la guía del Consorcio W3, un servicio web es un “conjunto de aplicaciones o de tecnologías con capacidad para interoperar en la Web. Estas aplicaciones o tecnologías intercambian datos entre sí con el objetivo de ofrecer unos servicios. Los proveedores ofrecen sus servicios como procedimientos remotos y los usuarios solicitan un servicio llaman-

do a estos procedimientos a través de la Web”².

En el contexto de la web, es la vía más extendida para explotar recursos entre ambientes, herramientas o aplicaciones. Permite desarrollar aplicaciones altamente especializadas en la gestión de una determinada tarea, sin sacrificar los niveles de especialidad o particularidad que a veces requieren los modelos de sistemas integrados. El modelo de desarrollo en base a servicios web se denomina Service Oriented Architecture (SOA).

Cada servicio web expone servicios y funcionalidades en una URL determinada, admite determinadas cláusulas y parámetros como argumentos válidos para brindar los servicios y, finalmente, brinda los servicios (o datos) de acuerdo a una estructura y formato determinados. A través de los tres aspectos antes mencionados es posible definir las características y capacidades de cada servicios web:

- Modalidad de acceso (URL, acceso abierto o restringido por clave, etc);
- Modelos de consulta;
- Modelos de respuesta.

En el caso particular del portal [gobiernolocal.gob.ar](http://www.gobiernolocal.gob.ar), la vía de acceso a los servicios web y por lo tanto a los datos es la

URL: <http://www.gobiernolocal.gob.ar/webservices/xml.php>

Al acceder a esa dirección sin ningún parámetro, se podrá ingresar a un ejemplo de datos básico de identificación de un municipio, en este caso, Roque Pérez.

Con respecto a las modalidades de consulta, el portal ofrece en la actualidad las siguientes modalidades:

- Datos de municipios;

2: W3C. Oficina española. Guía Breve de Servicios Web. Disponible en: <http://www.w3c.es/Divulgacion/GuiasBreves/ServiciosWeb> Fecha de consulta 2012-10-03

- Lista de indicadores estadísticos asociados a un municipio;
- Acceso cada uno de los indicadores asociados a un municipio;
- Lista de redes de municipios en las que participa un municipio;

De esta manera, es posible acceder a una infraestructura de servicios basada en datos en condiciones de uniformidad y persistencia, exponiendo datos codificados en condiciones óptimas de interoperabilidad y reuso.

Estamos explorando y construyendo una vía más en el camino de la consolidación de una capa de infraestructura de servicios públicos basados en datos en el ámbito del gobierno local en Argentina.

Invitamos e incitamos a todos los interesados en utilizar los datos disponibles en el portal gobiernolocal.gob.ar a avanzar en la agenda de reuso y explotación de datos abiertos. El portal ofrece una primera capa en términos de infraestructura, consideramos que los usos y servicios más significativos derivados de ella están aún por venir y serán el resultado de la intervención creativa y crítica de otros actores que diariamente participan del espacio común y público del gobierno local. Mientras seguimos trabajando en ampliar y mejorar la plataforma expuesta de datos, estamos abiertos y ansiosos a la expectativa de nuevos descubrimientos y realizaciones.

Referencias bibliográficas

DE SANTILLANA, GIOGIO (1966) El historiador y la teoría de la información. En: *El concepto de información en la ciencia contemporánea*. Siglo XXI: México DF, 1966

Principles of Open Government Data (2007) Disponible en: https://public.resource.org/8_principles.html Fecha de consulta 2012-10-03

RFC 1738 Uniform Resource Locators (URL) Disponible en: <http://tools.ietf.org/html/rfc1738> Fecha de consulta 2012-10-03

RFC 3986 Uniform Resource Identifier (URI): Generic Syntax. Disponible en: <http://tools.ietf.org/html/rfc3986> Fecha de consulta 2012-10-03

SWAN, ALMA (2012) *Policy Guidelines for development and promotion of Open Access*. UNESCO: París, 2012

W3C. Oficina española. *Guía Breve de Servicios Web*. Disponible en: <http://www.w3c.es/Divulgacion/GuiasBreves/ServiciosWeb> Fecha de consulta 2012-10-03

WRIGHT, ALEX (2007) *Glut: Mastering Information Through The Ages*. Joseph Henry Press, 2007