

# **Sugerencias a la norma PNE 178301 Ciudades Inteligentes. Datos abiertos (Open Data)**

## Ámbito

[Directiva europea del 2013](#)

[Norma Técnica de Interoperabilidad sobre reutilización](#)

[Raw data now](#)

[Cantidad de datasets y porcentajes](#)

[4.1.1.2. Liderazgo](#)

[4.4.1.3. Conjuntos de datos documentados](#)

[4.4.3.3. Datos documentados](#)

[4.4.3.6. Datos enlazados](#)

[4.5.1.2. Formato de los datos](#)

[4.5.2. Dimensión participación y colaboración](#)

## **Ámbito**

En el punto 4.1.1.1. se habla de la “existencia de una estrategia política de apertura de datos de la Entidad, incluidas Entidades dependientes, empresas públicas, etc”.

Si esta norma pretende ser de aplicación a cualesquiera niveles de la Administración, puede ser un problema definir el ámbito de aplicación. No es fácil acotar el sector público en torno a la Administración que se está acreditando, ya que incluye desde entes instrumentales hasta sociedades de participación múltiple.

Posibles soluciones:

- Que la propia Entidad defina el ámbito de aplicación de su servicio de datos abiertos y que la certificación recoja ese ámbito.
- Acotar la definición. P.e.: “todas aquellas organizaciones de las que se sea el titular mayoritario”.

Una idea complementaria es la de poner como requisito el de dar de alta a la Entidad y sus dependientes en Directorio Común de Unidades Orgánicas y Oficinas, teniendo en cuenta que:

- Es un trámite obligatorio para cumplir con el Esquema Nacional de Interoperabilidad.
- Es interesante para que funcionen bien las fichas en RDF ya que este sería el identificador para alguno de los metadatos (p.e.: “Organismo que expone y publica los datos”)

## **Directiva europea del 2013**

Existe una nueva directiva europea, que será transpuesta en un futuro cercano. Lo más relevante de esta norma es que eleva el grado de exigencia, hasta el punto de que inaugura un nuevo derecho para la ciudadanía: el de la reutilización de la información del sector público (Neelie Kroes dixit). Por otra parte, algunas normas como el RD y la reciente Ley de Administración Pública Vasca caminan también en esta línea de máxima exigencia.

Sugerimos tomar en cuenta esta perspectiva desde ahora, en el caso de que aún no se haya hecho.

## **Norma Técnica de Interoperabilidad sobre reutilización**

Algunas Administraciones (como mínimo, Open Data Aragón) han sometido el servicio de open data a una evaluación externa acerca del cumplimiento de la NTI de interoperabilidad.

Entendemos que podría establecerse una relación entre el cumplimiento de esta NTI y los criterios de la norma, de manera que una evaluación del la NTI podría servir para acreditar ciertos niveles, sobre todo en el dominio técnico de esta norma.

## **Raw data now**

Uno de los acuerdos más generalizados acerca de cómo abordar una estrategia de apertura de datos es el llamado “raw data now”.

Detrás de esta expresión hay dos ideas principales:

- “Release early, release often”: liberar el mayor número de datasets, en formatos estándar que puedan procesar las máquinas, sin esperar a mayor sofisticación tecnológica, para contar con muchos datasets liberados lo antes posible..
- Modelo centrado en la reutilización: en el diseño del servicio, priorizar todo aquello que favorezca la reutilización, por delante de otras consideraciones tecnológicas y organizativas.

El punto 4.4.3.1. de la norma desincentiva el raw data now, al exigir una mayoría (idealmente >90%) de datos primarios. Para puntuar más en este punto, es prudente no publicar datasets que se pueden liberar con poco esfuerzo, pero que son estadísticas u otras agregaciones que no pueden considerarse como datos primarios. Por lo tanto, va en contra de liberar esos datos. Más aún, castiga la publicación de datasets complementarios a un dataset primario y que pueden ser de interés para comunidades de reutilizadores concretas.

**Possible solución:** establecer una lista común de fuentes prioritarias y evaluar cuántas de estas se ofrecen como datos primarios, como sistema para evaluar hasta qué punto los datos son primarios. Además, valorar positivamente la publicación de otros datasets no primarios.

## Cantidad de datasets y porcentajes

El número de dataset publicados e inventariados son dos cifras clave, no sólo para el punto 4.5.1.1. (cantidad de datos publicados), sino como base para el cálculo de varios otros puntos. Surgen estas dificultades:

- Es muy difícil determinar qué es un dataset
- Los datasets son enormemente heterogéneos en tamaño y relevancia
- No hay un horizonte claro sobre lo que debería publicarse

Una tabla de 20 registros pesa tanto como una base de datos compleja que se ofrezca de manera primaria, en un solo dataset.

Entre los efectos esperables están:

- Declarar un inventario corto, para que el porcentaje de datasets publicados sea alto
- Multiplicar los datasets sin necesidad, desagregando tablas de una base de datos

**Possible solución:** la misma que en el punto anterior. Partir de una lista común de datasets relevantes y evaluar cuántos de estos se han publicado.

### 4.1.1.2. Liderazgo

Difícil evidenciar el nivel que se ha alcanzado. Tal vez podría ligarse a evidencias en cuanto a intervenciones públicas en materia de open data efectuadas por el líder político y a compromisos escritos existentes.

### 4.4.1.3. Conjuntos de datos documentados

Para alcanzar el 100% se establece este requisito: “Todos los conjuntos de datos cuentan con el 100% de los metadatos obligatorios y con el 100% de los opcionales”.

Parece excesivo, porque no queda claro que sea positivo en todos los casos publicar el 100% de metadatos del estándar DCAT. De hecho, se ha publicado en la plataforma Joinup un perfil de aplicación específico europeo (DCAT-AP), en su versión 1.0 del 2 de septiembre del 2013 [[https://joinup.ec.europa.eu/asset/dcat\\_application\\_profile/](https://joinup.ec.europa.eu/asset/dcat_application_profile/)], que recomienda algo diferente. Distingue entre:

- Metadatos obligatorios
- Metadatos recomendables
- Metadatos opcionales

Parece lógico que el nivel 3 se alcance con estos requisitos:

- Publicar todos los metadatos de catálogo recogidos en DCAT

- Publicar, respecto a los metadatos de los datasets:
  - Los declarados obligatorios en DCAT-AP
  - Los declarados recomendables en DCAT-AP
  - Una selección, superior al 50%, de los declarados opcionales en DCAT-AP, en función de su relevancia para la Entidad

#### 4.4.3.3. Datos documentados

Para empezar, el título despista un poco, ya que se parece mucho al 4.4.1.3. Sugerimos cambiar el 4.4.1.3. por “Metadatos de los conjuntos de datos”.

Hay, además, una errata en el nivel 0 de este punto, que dice:

*“Menos del 90% de los datos no están documentados.”*

Obviamente, hay que quitar el “no”. Y, en aras de la claridad, es mejor dar la vuelta a la frase, para que diga:

*“Están documentados menos del 90% de los datos”*

#### 4.4.3.6. Datos enlazados

El problema con este punto es que no hace ninguna referencia a la cantidad. En puridad, bastaría con tener un dataset como Linked Data (p.e., el dataset del propio catálogo) para puntuar el máximo.

Este ha sido un problema reiterado en referencia al modelo de 5 estrellas de Berners-Lee, que hay que corregir. Se suelen puntuar portales enteros de open data por la mera presencia de un pequeño subconjunto de datos en Linked Data, incluso si existen datasets que no pasan de la 2<sup>a</sup> estrella.

Sería de utilidad una orientación más clara de qué datasets es aconsejable ofrecer en Linked Data.

#### 4.5.1.2. Formato de los datos

El nivel 3 de este punto se define así: “El 90% de los conjuntos de datos se ofrecen en formatos estándares, estructurados, abiertos y que permite la identificación única de los recursos.”

En la nota a pie de página, se aclara que se refiere a:

*“Tipo RDF, etc. Se corresponde con el nivel de cuatro estrellas de la clasificación de 5 estrellas del Open Linked Data de Tim Berners-Lee (<http://5stardata.info/>)”*

Bien, tal como se dice en esa página, la 4<sup>a</sup> estrella no precisa necesariamente de publicar los datos en RDF, sino que la clave está en “use URIs to denote things, so that people can point at your stuff”. Incluso advierte de que “Understanding the structure of an RDF “Graph” of data can be more effort

than tabular (Excel/CSV) or tree (XML/JSON) data", lo que podría alzar una barrera para la reutilización por parte de muchas comunidades.

Por lo tanto sería interesante reforzar la idea de que la clave está en las URIs, más que en RDF.

Otra sugerencia respecto a los formatos tiene que ver con el hecho de que no se valora positivamente la oferta de más un formato. Tal como hemos visto en el caso del rdf, es interesante ofrecer formatos alternativos que coincidan con las necesidades diversas de grupos diferentes. Debería valorarse.

#### **4.5.2. Dimensión participación y colaboración**

En general, esta dimensión parece estar trabajada de forma más tentativa que las dimensiones técnicas.

En concreto, se echa en falta reclamar una coherencia con el aspecto "open" de la iniciativa. Esto es, debería medirse que la estrategia de datos abiertos tiene un carácter eminentemente abierto en todos sus aspectos, con acceso libre a sus elementos de gestión, a todos los documentos de gestión. Debe valorarse además, la incorporación de mecanismos de innovación abierta, dirigidos tanto hacia la sociedad como, de manera interna, hacia los propios trabajadores.

También podrían definirse mejor los canales de transparencia y participación, a partir de las mejores prácticas conocidas.