

دليل بيانات دبي

الأشكال

الإصدار 1.2 (ديسمبر 2020)



دبي الذكية
SMART DUBAI

الأشكال

<p>توضح هذه الوثيقة الحد الأدنى من المتطلبات الإلزامية وتوصيات لأفضل الممارسات فيما يتعلق بالأشكال التي ستستخدم عند نشر بيانات دبي المفتوحة والمشاركة، وتعد ذات أهمية خاصة في ضمان سهولة الوصول إلى البيانات واستخدامها من قبل مستخدميها، وقابلية التشغيل البيئي اللاحق مع البيانات الأخرى.</p>	<p>الغرض:</p>
<p>أخصائيو البيانات، الذين يخضعون لإشراف إداري البيانات مباشرة.</p>	<p>على من تقع المسؤولية الرئيسية عن تنفيذ هذا المعيار في كل جهة حكومية:</p>
<p>يستخدم معيار بيانات دبي هذا بعد إنشاء القائمة الأولية لجرد بيانات الجهة والانتهاء من تحديد الأولويات، كجزء من عملية الفهرسة وينبغي القيام بذلك بالتزامن مع تنفيذ الإجراءات المتعلقة بالبينات الوصفية و جودة البينات، وقبل تحميل البيانات على منصة دبي بالس.</p>	<p>متى تستخدم هذه الوحدة من دليل بيانات دبي</p>
<p>مؤسسة بيانات دبي</p>	<p>مالك الوثيقة:</p>
<p>يكون للمصطلحات الواردة في هذا المعيار نفس التعاريف الواردة في قاموس مصطلحات دليل بيانات دبي.</p>	<p>التعاريف:</p>
<p>نشرت هذه الوثيقة تحت شروط رخصة المشاع الإبداعي الدولية 4.0 من أجل تسهيل إعادة استخدامها من قبل الحكومات الأخرى والجهات الخاصة. وبإيجاز، هذا يعني أنه سيكون لكم مطلق الحرية في مشاركة المواد والتعديل عليها، بما في ذلك مشاركتها والتعديل عليها لأغراض تجارية، شريطة الإشارة لمؤسسة بيانات دبي بصفتها المالك لها من دون الإيحاء بكونها تجيز استخدامكم لهذه المواد.</p>	<p>الترخيص:</p>
<p>الإصدار 1.2 ، الصادر بتاريخ 9 ديسمبر 2020.</p>	<p>تاريخ الإصدار:</p>

نظرة عامة

يؤثر الشكل الذي تنشر به البيانات بصورة كبيرة على مدى سهولة العثور على تلك البيانات والاستفادة منها من قبل مستخدمي البيانات.

شكل البيانات هو الطريقة القياسية التي يتم ترميز المعلومات من خلالها لغرض تخزينها ونقلها عن طريق أجهزة الحاسب الآلي. ويحدد الشكل الطريقة التي يتم بها ترتيب البيانات بطريقة تسمح بقراءة البيانات بواسطة تطبيقات البرمجيات. يتم اختيار شكل البيانات على أساس فهمنا لنوع البيانات التي نتعامل معها في أي مجموعة بيانات محددة.

إن تحديد الشكل الصحيح لإتاحة أو مشاركة مجموعة بيانات هو جزء لا يتجزأ من [عملية الفهرسة](#) الأوسع لبيانات دبي. وتوضح هذه الوحدة المتطلبات التفصيلية للخطوة 3 من تلك العملية، وتهدف أيضاً إلى مساعدة الجهات الحكومية في نشر البيانات في أشكال مناسبة من خلال تحديد:

- [متطلبات الامتثال](#) لهذا المعيار
- [الإجراءات](#) التي يتعين عليكم اتباعها لاختيار الشكل المناسب
- [أفضل الممارسات](#) للأشكال الشائعة
- [كيفية تجنب الأخطاء الشائعة](#) عند اختيار أشكال البيانات
- كيف يمكن لجهتكم إثبات [امتثالها](#) لمتطلبات هذه الوحدة من دليل بيانات دبي.

المتطلبات

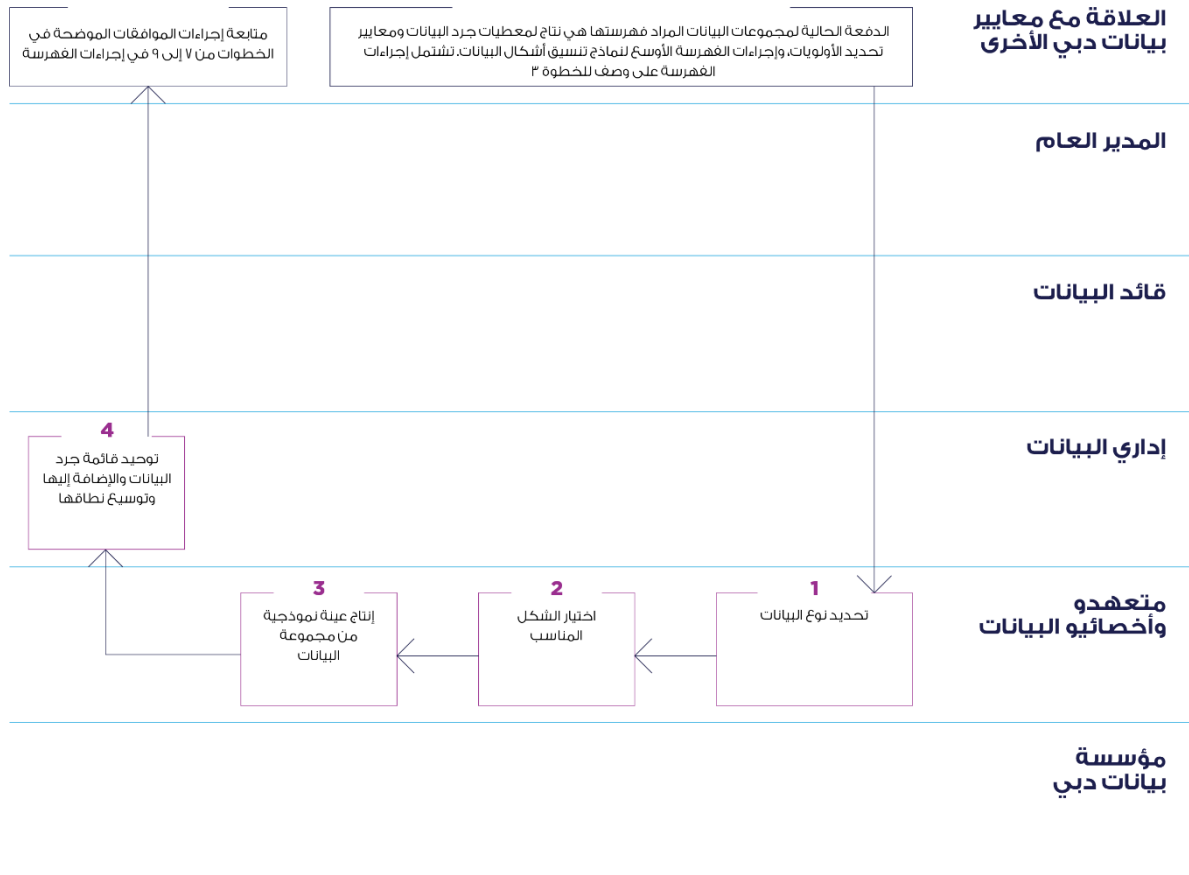
يتعين القيام بما يلي:

- 1- توفير جميع مجموعات البيانات التي سيتم نشرها أو مشاركتها على منصة دبي بالس في أشكال مناسبة ومنظمة وقابلة للقراءة آلياً ومفتوحة كما هو مبين في معيار بيانات دبي هذا، ما لم:
 - أ- تشترط مؤسسة بيانات دبي نشر البيانات باستخدام معايير مفتوحة دولية ذات نطاق خاص (مثل معيار تغذية حركة النقل العام لنقل لبيانات)؛ أو
 - ب- يتم الاتفاق على استثناء معين مع مؤسسة بيانات دبي بسبب وجود حالة عمل واضحة أو لوجود سبب عملي يحول دون اتباع هذا المعيار.
 - 2- بالنسبة للبيانات التي تتوفر لها رموز تعريفية فريدة (أي، تم تمييز الرموز التعريفية الفريدة للسجلات أو العناصر التي تتضمنها البيانات)، يتعين استخدام هذه الرموز التعريفية الفريدة وإدراجها في البيانات. ومن المستحسن:
- 1- أن تنشر الجهات أيضاً بيانات جدولية في أداة جدول بيانات تحليلية واحدة (مثل جداول بيانات إكسيل أو الوثائق المفتوحة) تحتوي على كافة البيانات وجميع البيانات الوصفية القابلة للقراءة آلياً.
 - 2- أن يتضمن كل سجل ضمن مجموعة بيانات رمزاً تعريفياً فريداً لموضوع هذا السجل.

¹ تشير المتطلبات والمعايير والإجراءات الموصوفة هنا فقط إلى البيانات المنظمة. وعلى الرغم من أن برنامج بيانات دبي يشمل أيضاً على المدى الطويل بيانات غير منظمة، إلا أن تركيز المبادرات وبيانات دبي ينصب حالياً على البيانات المنظمة فقط (لمزيد من المعلومات انظر الوحدة 1 – النظرة العامة)

الإجراءات

يوضح القسم أدناه الإجراءات التي يتعين اتباعها عند تحديد أشكال البيانات، وسيتم بعد ذلك وصف كل خطوة بمزيد من التفصيل.



تحديد نوع مجموعة البيانات

الخطوة الأولى لاختيار شكل البيانات هي تحديد نوع البيانات التي تتعامل معها، ذلك أن الأنواع المختلفة من البيانات لها خصائص مختلفة وتحتاج إلى تنسيقها بطرق مختلفة.

البيانات الجدولية

معظم البيانات الحكومية هي بيانات جدولية، فإذا كانت البيانات التي تتعامل معها عبارة عن قائمة أو كان من المنطقي تسجيلها في جدول بيانات فمن المؤكد أنها بيانات جدولية.

تتألف البيانات الجدولية من صفوف يشكل كل منها سجلاً فردياً في مجموعة البيانات، ومن أعمدة، يشكل كل منها حقل واحد من السجل. على سبيل المثال، قد تظهر مجموعة بيانات عن المدارس بالشكل التالي:

الرمز التعريفي المميز	الاسم	أعلى سن	أدنى سن
AB292	بلو ووتر هاي	11	5
HG383	أكاديمية جرين تري	11	5

الصف الثاني الذي يحتوي على عبارة "بلو ووتر هاي" هو عبارة عن قيد حول بلو ووتر هاي، ويحتوي العمود المعنون "أعلى سن" على البيانات المتضمنة في حقل أعلى سن في كل مدرسة.

البيانات الجغرافية المكانية

تتعلق البيانات الجغرافية المكانية بمعلومات حول كيفية رسم الأشياء على الخريطة، ويمكن تصنيف البيانات على أنها بيانات جغرافية مكانية عندما:

- تحتوي على إحداثيات مستخدمة للإشارة إلى شيء على الخريطة - مثل زوج خطوط الطول والعرض - على سبيل المثال مواقع أماكن وقوف السيارات أو المكتبات العامة.
- تحتوي على الشكل الذي سوف نرسمه على خريطة ما للإشارة إلى منطقة معينة. على سبيل المثال، بيانات حول منطقة متجمعات المياه للمدرسة على سبيل المثال: الهيئات الإدارية للمناطق التعليمية، أو أماكن تقسيم المناطق للحصول على تصاريح التخطيط.

بيانات لحظية

يتم توفير بيانات لحظية بصورة فورية عن طريق واجهة برمجة التطبيقات التي يمكن استيعابها في تطبيقات البرمجيات الأخرى. يطلق وصف بيانات الوقت الحقيقي على البيانات إذا كانت تتغير بصورة متواترة لدرجة أن معظم الأسئلة التي قد تطرحها عن هذه البيانات سوف تصبح غير صالحة سريعاً.

ومن الأمثلة على ذلك حالة القطارات على شبكة السكك الحديدية، أو معلومات عن الرحلات الجوية الحالية - المغادرين والقادمين والتأخيرات في المطارات.

البيانات المنظمة في شكل غير جدولي

قد تكون بعض البيانات مهيكلة، لكنها لا تصلح لأن تكون في شكل جداول بطريقة طبيعية. فإذا كانت بياناتكم هرمية أو تحتوي على العديد من المستويات، فهي على الأرجح بيانات مهيكلة في شكل غير جدولي. ومن الأمثلة على ذلك المخطط التنظيمي لإدارتك أو خطة المشروع.

اختر شكلاً مناسباً للبيانات

بعد تحديد نوع البيانات التي تتعامل معها، يمكننا استخدام الجدول التالي لتحديد الشكل المطلوب استخدامه:

نوع البيانات	الشكل المطلوب	الشكل الموصى به
البيانات الجدولية	ملف قيم مفصولة بفاصلة (CSV)	ملف قيم مفصولة بفاصلة (CSV) وملف إكسيل مع التعاريف والملاحظات
البيانات الجغرافية المكانية	<ul style="list-style-type: none"> • مجرد إحداثيات ومعلومات حول هذه النقطة • أشكال المناطق 	ملف قيم مفصولة بفاصلة (CSV) GeoJSON or KML
بيانات الوقت الحقيقي	عبر واجهة برمجة التطبيقات	GeoJSON or KML
مهيكلة في شكل غير جدولي	شكل مناسب قابل للقراءة آلياً، يتوافق مع معيار مفتوح حيثما كان ذلك متاحاً. على سبيل المثال، JSON, XML, TTL, RDF, GTFS	

الأشكال البديلة

إذا كنت ترغب في استخدام شكل بديل غير المذكورة أعلاه، فأنت بحاجة إلى إنشاء حالة عمل تضم التفاصيل الآتية:

- الشكل الذي اخترته
- وصف مجموعة البيانات ذات الصلة
- سبب ملائمة هذا الشكل لمستخدمي مجموعة البيانات هذه
- التأكيد على أنك قد بحثت عن الشكل الأكثر انفتاحاً وتوحيداً وقابلية للقراءة الآلية والمناخ كذلك لتلبية احتياجاتك، وأن هذا الشكل يتوافق مع التوجيهات المتعلقة اختيار معايير مفتوحة ومقبولة وشفافة كما هو منصوص عليه في الـ 6 و 7 من دليل بيانات دبي حول [كيفية إدارة معايير بيانات دبي](#).

وينبغي الاتفاق على هذا الشكل البديل مع إداري البيانات أو قائد البيانات (حسبما هو معمول به في جهتك)، ومن ثم يتم تقديم الشكل المختار إلى فريق مؤسسة بيانات دبي للموافقة عليه.

إنتاج عينة مجموعة البيانات

بمجرد تحديد الشكل المناسب، يتم الانتقال إلى المرحلة التالية المتمثلة في إنتاج عينة مجموعة البيانات.

قد تكون عينة مجموعة البيانات على هيئة ملف أو مجموعة ملفات تمثل ما قمت بنشره في كل مرة يتم فيها تحديث مجموعة البيانات. ويمكن إنتاج عينة مجموعة البيانات على هيئة ملف كامل للبيانات أو بعض البيانات. إذا كان يجري تحديث بياناتكم بشكل مستمر، فإن عينة مجموعة البيانات يمكن أن تكون شريحة معينة من البيانات.

تمثل العينة ما سيتم تحميله في المنصة وإتاحته للمستخدمين. الغرض من إنتاج هذه العينة هو مساعدة متعهد البيانات وإداري البيانات وقائد البيانات ومؤسسة بيانات دبي في التحقق من أن البيانات:

- تتوافق مع معايير الفهرسة
- تبدو منطقية كمجموعة بيانات

توحيد قائمة الجرد والإضافة إليها وتوسيع نطاقها، والتماس الموافقة على هذه الإجراءات باعتبارها جزءاً من عملية فهرسة البيانات

بمجرد الانتهاء من [تصنيف](#) كل مجموعات البيانات التي سيتم [تحميلها](#) ونشرها في الدفعة الحالية وبعد اختيار [الأشكال](#) المناسبة لها والانتها من الإجراءات المتعلقة [جودة البيانات](#)، يقوم إداري البيانات بتضمين كافة النتائج والبيانات [الوصفية](#) لمجموعات البيانات هذه في قائمة جرد بيانات الجهة.

وينبغي بعد ذلك أن يجري قائد البيانات مراجعة داخلية لقائمة الجرد المفهرسة الناتجة بعد تضمين النتائج والبيانات الوصفية، ومن ثم يتم إرسالها إلى مؤسسة بيانات دبي كجزء من العملية الأوسع الموصوفة في [إجراءات الفهرسة](#) (يرجى الرجوع على وجه الخصوص إلى الخطوات 7 و 8 و 9 من هذه الوحدة). وستقوم مؤسسة بيانات دبي بمراجعة البيانات المفهرسة باتباع الإجراءات الموضحة في [وحدة معايير قبول نشر البيانات](#) من دليل دبي.

أفضل الممارسات للأشكال الشائعة

تتعلق المتطلبات التالية بكيفية التأكد من جودة البيانات حسب نوع البيانات.

وعموماً، يعتبر شكل مجموعة البيانات مطابقاً لإرشادات جودة البيانات الموصى بها إذا كان:

- لا يستخدم شكل المستندات المنقولة (PDF) لنشر البيانات المنظمة
 - يتضمن الرمز التعريفي الفريد لموضوع كل سجل في كافة السجلات التي يوجد بها هذا الرمز التعريفي
 - الشكل الذي تم اختياره مناسب لنوع البيانات أو في معيار محدد النطاق متفق عليه.
- بالإضافة إلى المعايير العامة المذكورة أعلاه، هناك معايير محددة للشكل كما هو مبين أدناه.

البيانات المدرجة في ملف قيم مفصولة بفاصلة (CSV)

يعتبر شكل مجموعة البيانات المدرجة في ملف قيم مفصولة بفاصلة (CSV) مطابقاً بشرط:

- أن يحتوي على صف العنوان الذي يتضمن اسم العمود
- أن تكون تنسيقات التواريخ أو الأرقام متسقة في كامل الملف
- ألا يحتوي على صفوف فارغة
- ألا يحتوي على صفوف ذات خلايا مفقودة أو إضافية
- ألا يستخدم أسماء العناوين أكثر من مرة في نفس الملف
- ألا يتضمن أي تعليق أو نص توضيحي

البيانات المنظمة في شكل غير جدولي

يعتبر شكل البيانات المنظمة في شكل غير جدولي مطابقاً بشرط:

- أن يكون متوافقاً مع معيار مفتوح موجود مسبقاً لعرض هذه البيانات، مثل معيار تغذية حركة النقل العام لنقل البيانات (GTFS)، بوبولو (Popolo)، ومعيار نشر المهام (Schema.org)؛ أو
- أن يكون مدرجاً ضمن معيار مفتوح صالح للقراءة آلياً مثل JSON, XML, TTL
- أن تكون بنية هذه البيانات موثقة بشكل واضح ومنشورة إلى جانبها
- أن تكون بنية البيانات مناسبة لإعادة الاستخدام بالنظر إلى طبيعة المجال الذي تتصل به البيانات.

البيانات الجغرافية المكانية

يعتبر شكل مجموعة بيانات جغرافية مكانية مطابقاً بشرط:

- أن يكون قد تم نشره في معيار صالح لجيوجسون الجغرافي (GeoJSON) أو كيه إم إل (KML) (انظر المزيد من النصائح حول الأشكال المحددة لأدوات التحقق).

كيفية تجنب الأخطاء الشائعة عند اختيار

أشكال البيانات

ملفات القيم المفصولة بفاصلة (CSV)

تتوفر أدوات لنت (البرامج التي تحلل التعليمات البرمجية والبيانات وغيرها من الأخطاء المحتملة) مثل <http://csvlint.io/> للبيانات المدرجة في ملفات القيم المفصولة بفاصلة ويمكن استخدامها لتحديد الأخطاء في ملفاتكم في وقت مبكر.

إذا كان حقل ما يحتوي على فاصلة أو سطر نهاية أو اقتباس مزدوج، يتم التخلص من الحقل عن طريق ربطه مع علامات الاقتباس المزدوجة، وتتم مضاعفة أي علامات اقتباس مزدوجة في أي حقل يتم التخلص منه على النحو المذكور.

بيانات الجغرافية المكانية

تتوفر أدوات لنت لمعيار جيوجسون (GeoJson) مثل <http://geojsonlint.com/> والتي ستساعد في التعرف على الأخطاء في ملفات بياناتكم. بالنسبة لمعيار كيه ام ال (KML). وللتحقق من توافق بياناتكم مع معيار كيه ام ال سكيما (KML Schema) يمكن الدخول إلى <https://developers.google.com/kml/schema/kml21.xsd?csw=1>

تتوفر أيضاً أدوات مفتوحة ذات جودة عالية لتحويل البيانات الجغرافية المكانية إلى أشكال، مختلفة ويمكن إدراجها في الخطوط المؤتمتة لتوليد مجموعات البيانات لنشرها بسهولة في أشكال متعددة.

ومن الأمثلة الجيدة على ذلك <http://www.gdal.org/ogr2ogr.html>

البيانات المنظمة في شكل غير جدولي

بالنسبة للعديد من أنواع مجموعات البيانات الشائعة، توجد معايير مفتوحة لعرض تلك المعلومات كبيانات مهيكلية وينبغي إعادة استخدامها بقدر الإمكان.

ومن أمثلة هذه المعايير ما يلي:

- معايير سكيما المتاحة على <http://schema.org/>
- معيار بوبولو (Popolo) للبيانات للأشخاص والمنظمات والتصويت <http://www.popoloproject.com/>

وبصفة عامة، يجب أن تستخدم البيانات المنظمة في شكل غير جدولي معيار جسون (JSON)، ما لم يكن هناك سبب واضح لاستخدام شكل بديل، مثل معيار مشترك في شكل بديل (مثل معيار تغذية حركة النقل العام لنقل البيانات GTFS).

الإمتثال

يجب على إداري البيانات أن يراجع كافة مجموعات البيانات النموذجية لضمان ما يلي:

- أن الشكل الذي يتم اختياره مناسب [لنوع البيانات](#) كما هو مبين في هذا المعيار أو في شكل موحد ومحدد النطاق (مصادق عليه من قبل مؤسسة بيانات دبي).
- بالنسبة إلى البيانات التي تتوفر لها رموز تعريفية بالفعل، أن يحتوي كل سجل على رمزه التعريفي الفريد.
- اتباع أفضل الممارسات فيما يتعلق ببيانات السجل المرجعي الأساسي ومجموعات البيانات ذات القيمة العالية.