

دليل بيانات دبي

جودة البيانات

الإصدار 1.2 (نوفمبر 2020)



دبي الذكية
SMART DUBAI

جودة البيانات

<p>تصف هذه الوثيقة إرشادات جودة البيانات الموصى بها، والتي تهدف إلى تحسين إمكانية إعادة استخدام بيانات دبي وموثوقيتها. فلا تحدد هذه الوثيقة متطلبات البيانات الإلزامية، ولكن ينبغي أن تستخدمها الجهات للمساعدة في تحسين جودة البيانات مع مرور الوقت لبياناتها الأكثر قيمة.</p>	<p>الهدف</p>
<p>متعهد البيانات وأخصائي البيانات.</p>	<p>المسؤولية الأساسية عن تنفيذ هذا النموذج في كل مؤسسة حكومية</p>
<p>أثناء عملية فهرسة البيانات.</p>	<p>متى يمكن استخدام نموذج دليل بيانات دبي الحالي</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>التعريفات</p>
<p>نشرت هذه الوثيقة بموجب شروط الرخصة الدولية للمشاع الإبداعي 4.0 لتيسير إعادة استخدامها من قبل الحكومات الأخرى ومؤسسات القطاع الخاص. وبإيجاز، يعني ذلك أن لك مطلق الحرية في اختيار مشاركة المواد والتعديل عليها، بما في ذلك مشاركتها والتعديل عليها للأغراض التجارية، شريطة أن تقدم الضمان المناسب لمؤسسة بيانات دبي كمالك لها وألا تفترض أن مؤسسة بيانات دبي ستجيز استخدامك لهذه المواد.</p>	<p>الترخيص</p>
<p>الإصدار 1.1، الصادر بتاريخ 23 أبريل 2018.</p>	<p>الإصدار</p>

نظرة عامة

تقدم هذه الوثيقة نظرة عامة حول الاثني عشر بعداً لجودة البيانات المتأصلة في أفضل الممارسات الدولية والتي من خلالها يجب على الجهات الحكومية أن تسعى مع مرور الوقت إلى تحسين البيانات التي تديرها. وتحدد الوثيقة ما يلي:

- [متطلبات الامتثال لهذا المعيار](#)
- [السبب وراء أهمية جودة البيانات](#)
- [الاثني عشر بعداً لجودة بيانات دبي](#)
- [العملية اللازمة لتلبية هذا المعيار](#)
- تفاصيل [تقييم جودة البيانات](#) التي ينبغي أن يقوم بها متعهدي البيانات لأصول البيانات ذات الأولوية
- [الإرشادات التوجيهية](#) الموصى بها حول جودة البيانات
- كيف يمكن للجهة التابعة لك أن تثبت [الامتثال](#) مع هذا النموذج من دليل بيانات دبي.

المتطلبات

تعتبر هذه الوثيقة دليل للممارسات الجيدة؛ ولا يوجد فيها شرط إلزامي يجب أن تفي به بيانات دبي. يتم تضمين معايير جودة البيانات الوحيدة التي يجب على الجهات الوفاء بها في معايير بيانات دبي المنفصلة المتعلقة بالنسبة للأشكال والبيانات الوصفية، أو يتم معالجتها تلقائياً من خلال منصة دبي بالس. ومن ثم، لا تحتوي هذه الوثيقة على أي متطلبات إلزامية إضافية. ويرجع ذلك إلى أنه في حين يتم تشجيع الجهات على تبني ممارسات جيدة، يجب ألا يشكل ذلك عائقاً أمام نشر البيانات.

ومع ذلك، يجب على متعهدي البيانات القيام بما يلي:

- 1- مراجعة [الإرشادات التوجيهية](#) الموصى بها حول جودة البيانات الواردة في هذه الوثيقة
- 2- وضع تقرير لجودة البيانات من أجل تلخيص الوضع الحالي للجودة وسرد القيود والمشاكل المعروفة في مجموعة البيانات التي يقومون بفهرستها.

لذا، يوصى بما يلي:

- 1- أن تتوافق مجموعات البيانات مع جميع [الإرشادات التوجيهية](#) الموصى بها حول جودة البيانات
- أن تقوم الجهات (على مستوى الجهة أو الإدارة أو وحدة الأعمال أو مجموعة البيانات -أيهما أنسب) بوضع خطة لكيفية تحسين جودة البيانات وتلبية المزيد من [الإرشادات التوجيهية](#) الموصى بها حول جودة البيانات مع مرور الوقت لنسبة كبيرة من بياناتها - بدءاً من البيانات الأكثر قيمة والبيانات الأكثر استخداماً أو طلباً.

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 (المزيد من المعلومات انظر الوحدة 1 - النظرة العامة)

لماذا تعد جودة البيانات هامة

لتحقيق أهداف برنامج دائرة دبي الذكية وتأثيره المقصود، يجب أن يكون هناك إعادة استخدام قيمة لبيانات دبي وزيادة جودة الخدمات وعددها المقدمة للمواطنين والشركات في دبي. ولتحقيق ذلك، لا يكفي نشر البيانات فقط؛ بل يحتاج مستخدمو البيانات أيضاً إلى الثقة بأن البيانات ملائمة للغرض، وسيتم الحفاظ عليها باستمرار مع مرور الوقت. وإذا أدت البيانات إلى تغيير المخطط أو التفاصيل أو توقف نشرها على فترات منتظمة، فإنها تصبح أقل موثوقية. وهذا بدوره يجعل الاستثمار في بناء تطبيق مفيد أو عملية الإدخال أمر غير اقتصادي إذا زادت احتمالية اختفاء أو استمرار تغيير البيانات التي يعتمد عليها تطبيق أو خدمة. وهذا يعني أن الحكومة بحاجة إلى التغيير من مجرد كونها منتجة بيانات إلى جهة نشر بيانات – جهة نشر مع العملاء للقيام بالخدمة.

ويوضح هذا المعيار أفضل الممارسات الدولية لنشر البيانات على مستوى عالٍ من الجودة، لبناء الثقة وتشجيع الاستثمار في الخدمات والمنتجات والأبحاث النهائية.

مصفوفة جودة بيانات دبي

تلتزم حكومة دبي بتحسين جودة بياناتها بمرور الوقت من خلال تقييم البيانات مقابل كل من الأبعاد السبعة الملخصة في الجدول أدناه.

يجب اعتماد مستوى جودة البيانات مقابل مصفوفة جودة البيانات في الملحق أ ، والتي تحدد خمسة مستويات من النضج في سبعة أبعاد رئيسية لجودة البيانات.

أبعاد جودة البيانات هي:

أبعاد الجودة	وصف الملخص	المستوى المطلوب
1- الملكية والسلطة	<p>تتم إدارة البيانات بواسطة متعهد بيانات ، يكون مسؤولاً عن جودة البيانات والتأكد من أنها تلبي احتياجات المستخدم.</p> <p>البيانات التي يتم نشرها أو تبادلها ليست نسخة ويمكن الوصول إليها من المصدر. يقوم المستخدمون بالوصول إلى البيانات الأصلية وإعادة استخدامها، وبالتالي تقليل الازدواجية والأخطاء.</p> <p>تتم إدارة البيانات المرجعية الأساسية كسجل دبي مع مالك موثوق كما يتم استخدامها ضمن مجموعات البيانات الأخرى ذات الصلة.</p>	إلزامي
2- الوصول الى البيانات	<p>يسهل الوصول الى البيانات واستخدامها ، لأن:</p> <p>أ. لديها بيانات وصفية شاملة لتعزيز قابلية الاكتشاف مما يعني أنه يجب على الجهات نشر بيانات وصفية شاملة تصف بياناتها وتعطي سياقاً لها؛</p> <p>ب. يتم نشرها واستخدامها في تنسيقات مناسبة مفتوحة يمكن قراءتها آلياً (تقليل متطلبات شراء برمجيات احتكارية محددة والتأكد من أنها قابلة للتشغيل البيني مع البيانات الأخرى)</p> <p>ج. متاحة في مجموعة متنوعة من الأشكال ؛</p> <p>د. متاحة للتنزيل بالجملة أو عبر واجهة برمجة التطبيقات إما على الشبكة أو من خلال نظام أساسي مع وصول دائم موثوق به يتم دعمه بمرور الوقت.</p>	إلزامي: المتطلبات مفصلة في معيار البيانات الوصفية و معيار تنسيقات البيانات

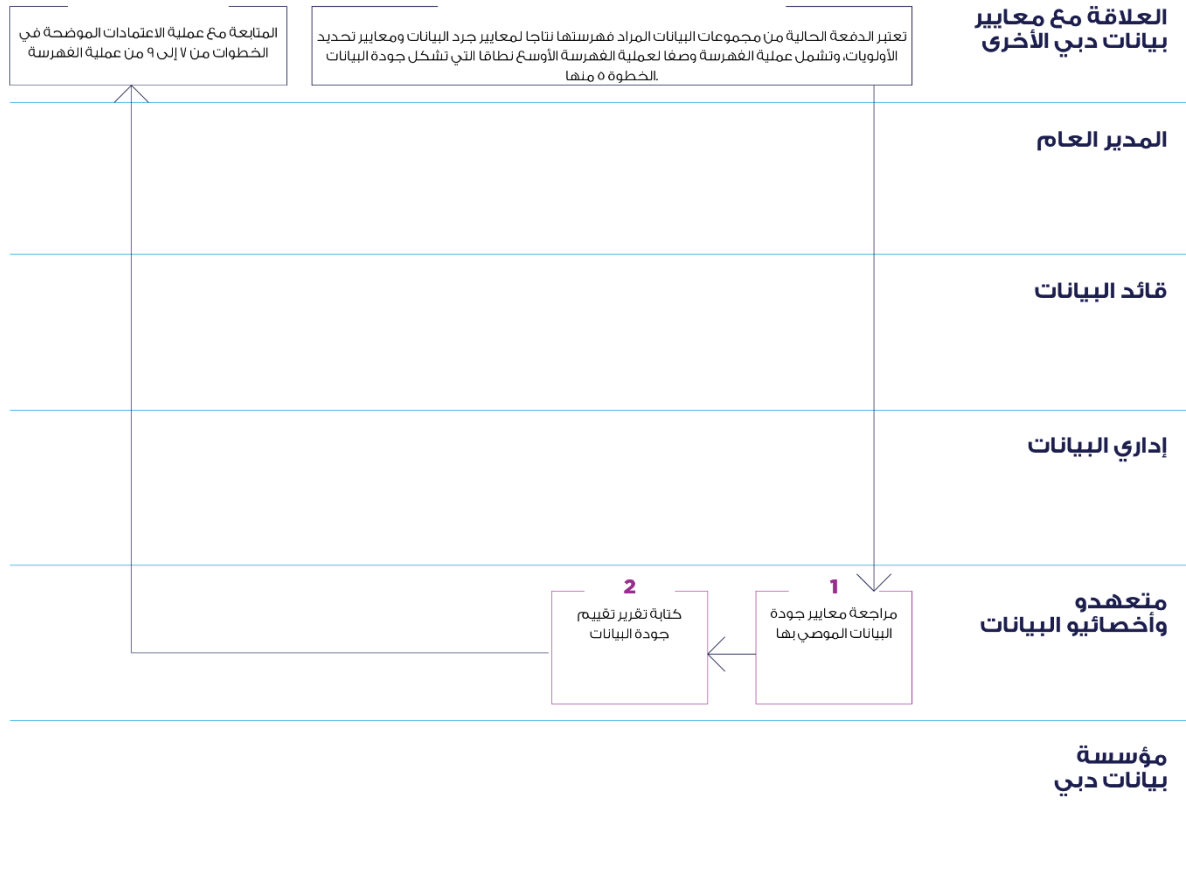
أبعاد الجودة	وصف الملخص	المستوى المطلوب
	<p>ه. إنشاء آليات وصول مستمرة ودائمة إلى البيانات، والتي تتضمن الالتزام بالحفاظ على إتاحة البيانات ، عبر الإنترنت وفي نفس الموقع لعدة سنوات ، ونشر تفاصيل حول سياسة الاحتفاظ بالبيانات التي سيتم تطبيقها.</p>	
3- الدقة	<p>البيانات دقيقة بما يكفي للاستخدام المقصود وأي ثغرات أو قيود معروفة أو تقديرات تقريبية أو أخطاء موصوفة موضحة حتى يفهم المستخدمون قيود البيانات. حقول البيانات الأساسية مثل التواريخ والمواقع ومعرفات الجهات يجب أن تستخدم التسمية المتعارف عليها.</p>	إلزامي
4- الوصف	<p>البيانات لها سياق لتسهيل التعريف للمستخدمين المحتملين بما هو موجود في البيانات ومدى موثوقيتها حتى يتمكنوا من الحكم بشكل فعال على ما إذا كانت مناسبة لغرضهم. هذا يعني أنه يجب أن تحتوي جميع مجموعات البيانات على بيانات وصفية مرتبطة ومخطط يحدد نطاقات وقيمة كل حقل. يجب أن يكون المستخدمون قادرون على فهم كيفية إنشاء البيانات ومعالجتها، والتغطية زمنية والجغرافية والتفصيل والقيود.</p>	إلزامي: يرجى الرجوع إلى المتطلبات مفصلة في معيار البيانات الوصفية و الإرشادات التوجيهية حول جودة البيانات أدناه
5- حداثة البيانات	<p>أ. يتم نشر البيانات أو إتاحتها في الوقت الفعلي أو بعد وقت قصير من إنشاء البيانات. يجب أن تكون البيانات التي يتم نشرها لإعادة استخدامها كبيانات مفتوحة أو يتم تبادلها مع جهات أخرى هي نفس البيانات المستخدمة للغرض المقصود منها داخل الجهة المُنتجة للبيانات.</p> <p>ب. يجب أن تلتزم الكيانات بجدول نشر واضح لعدد المرات التي سيتم فيها تحديث كل مجموعة بيانات، ومتى وبأي بيانات (على سبيل المثال، سننشر البيانات اعتباراً من 10 مايو)</p> <p>ج. التأكد من إصدار البيانات ونشرها على المنصة في الوقت المناسب – حتى يتم تحديث البيانات وإتاحة أحدث نسخة من البيانات.</p>	إلزامي: يرجى الرجوع إلى المتطلبات مفصلة في معيار البيانات الوصفية و الإرشادات التوجيهية حول جودة البيانات أدناه
6- الاكمال	<p>تعتبر البيانات منطقية باعتبارها مجموعة بيانات كاملة. يجب أن تكون قابلة للاستخدام دون الحاجة إلى بيانات أخرى لفهمها أو استخدامها. هذا يعني أنه يجب نشر البيانات أو تبادلها كمجموعات بيانات شاملة وبأعلى</p>	إلزامي

أبعاد الجودة	وصف الملخص	المستوى المطلوب
	مستوى من الدقة والتغطية المتاحة، ويجب وضع علامة على السجلات المفقودة ذات الصلة.	
7- التحقق من الصحة	<p>يجب أن تكون البيانات صحيحة، كما يجب بذل الجهد لضمان دقتها وموثوقيتها بمرور الوقت. هذا يعني بالنسبة للبيانات الأساسية والتي يتم تحديثها بشكل متكرر:</p> <p>أ. استخدام مخطط حيث يجب على الجهات نشر أو إنشاء مخططات موحدة للبيانات توضح بالتفصيل ما تمثله كل سمة والتحقق من الصحة / المفردات المستخدمة؛ وإعادة استخدام المخططات للبيانات المحدثة بانتظام.</p> <p>ب. وجود نموذج بيانات واضح مع معرفات فريدة وصفات الرئيسية في البيانات بحيث تكون هناك طريقة قياسية للإشارة إلى الصفات أو الكيانات المشتركة عبر جميع البيانات الحكومية لتمكين ربط البيانات معاً. بالنسبة للأشياء التي تظهر في سجلات دبي، يجب أن يكون المعرف الفريد هو نفسه الذي يتم الاحتفاظ به كمصدر موثوق هناك</p> <p>ج. تنقية البيانات واختبارها بانتظام لإزالة الأخطاء أو التكرارات.</p>	<p>إلزامي: يرجى الرجوع إلى الإرشادات التوجيهية حول جودة البيانات أدناه</p>

العملية

كجزء من عملية [الفهرسة](#) الأوسع نطاقاً "ولكل بيانات وصفية" يجب على متعهد البيانات المسؤول:

- 1- مراجعة [الإرشادات التوجيهية](#) الموصى بها حول جودة البيانات ووضع العمليات أو إضافة تفاصيل البيانات الوصفية ذات الصلة لتلك التي يمكن تلبيتها بسهولة
- 2- إجراء تقييم لجودة البيانات لتوفر وصفا موجزا لجودة البيانات الحالية الخاصة بالبيانات الوصفية وأية مشاكل وقيود معروفة تتعلق بالجودة (بالإشارة إلى المعايير الموصى بها) وإنشاء خطة، حسب الاختيار، لتلبية المزيد من المعايير في الوقت المناسب.



توصيات لمجموعات البيانات المحددة على أنها سجلات دبي:

- يجب على فريق مؤسسة بيانات دبي التحقق من التوافق والعلاقة بين السمات والمخططات للسجلات المماثلة؛ توحيد مصادر البيانات
- يجب أن تؤدي تعديلات وتحديثات سجلات دبي إلى إصدار تقرير خاص للموافقة السريعة وعملية التحميل
- يجب أن تشير سجلات دبي كجزء من وصف السمة إلى أي منها يعتبر إلزامياً لاعتبار السجل كاملاً. يمكن اعتبار السمات الأخرى اختيارية. بخلاف المجموعات البيانات الأخرى، يجب أن تسجل السجلات السمات الإلزامية بوضوح في تعريف البيانات الوصفية والمخطط

تقييم جودة البيانات وخطة التحسين

يلتزم متعهد البيانات بوضع تقرير لكل مجموعة بيانات يفصل ما يلي:

- حالة الجودة الحالية فيما يتعلق بالملكية والسلطة ، وإمكانية الوصول ، والدقة ، والوصف ، والتوقيت ، والاكتمال ، والتحقق
- أي ملاحظات عامة حول الجودة وخاصة حول المشاكل المعروفة أو القيود التي تتعلق بالجودة.

بشكل اختياري، يوصى متعهد البيانات و / أو فريق البيانات التابع للجهة بإعداد خطة لكيفية تحسين الإدارة ذات الصلة لمستوى الجودة هذا مع مرور الوقت. ويمكن أن يشمل ذلك:

- عمليات الأعمال الجديدة التي سيتم وضعها
- الطرق الجديدة للجمع أو المعالجة
- التدريب المخطط وتوحيد المعايير على مستوى الجهة
- الجداول الزمنية والبنود الرئيسية للتحسينات
- ما هو المستوى الجديد المتوقع للجودة
- الأدوار والموظفين المسؤولين عن هذه الخطة

الإرشادات الموصى بها بشأن جودة البيانات

يصف هذا القسم من الوثيقة الممارسات الجيدة في أبعاد جودة البيانات التي لم يتم تناولها بمزيد من التفصيل في معايير بيانات دبي المنفصلة حول [صيغ البيانات](#) و [البيانات الوصفية](#) أو معالجتها تلقائياً من خلال منصة دبي بالس. وهذه الأبعاد هي:

- [المخطط](#)
- [الرموز التعريفية الفريدة](#)
- [المصدر والأولية](#)
- [الجداول الزمنية](#)
- [الاستيفاء](#)

المخطط

يُوصى القائمين على نشر كل من بيانات دبي المفتوحة والمشاركة بنشر مخطط لمجموعات البيانات الخاصة بهم كجزء من البيانات الوصفية. ويعمل المخطط القابل للقراءة آلياً كضمان على صيغة مجموعة بيانات. وسيظهر المخطط بشكل طبيعي من عملية [أشكال البيانات](#)، وهو ببساطة وصف لهذا التنسيق بالإضافة إلى ضمان أن الإصدارات المستقبلية للبيانات ستستخدم نفس التنسيق.

كما يُتيح إنشاء أحد المخططات إمكانية التحقق التلقائي من مجموعات بيانات، فضلاً عن جعلها أسهل بكثير للأطراف الخارجية لبناء الأدوات حول مجموعات بيانات.

يجب أن تستخدم بياناتك، حيثما أمكن، أي مفردات قياسية دولية أو محلية (مثل رموز الدولة "ISO-3166-2-alpha" أو سجل أساسي للجهات الحكومية)، ثم تشير إليها في المخطط.

الخطوات:

1- هل تستخدم مخططاً بالفعل؟ بشكل عام، يحتوي امتداد "KML" أو التنسيقات محددة القطاع مثل "GTFS" بالفعل على مخطط محدد خارجياً تقوم باستخدامه. وإذا كان الأمر كذلك، فما عليك سوى ربط ذلك بحقل مخطط.

2- إذا لم يكن لديك مخطط حالياً وبياناتك عبارة عن ملفات بصيغة "CSV" أو "JSON" المنظمة، فحينئذ:

- أ اتبع معيار [شكل البيانات](#) لإنتاج ملف منسق بشكل جيد (باتباع إرشادات أفضل الممارسات)
 - ب قم بتوثيق الحقول باستخدام المخطط المناسب (المدرج أدناه).
 - 3- قم بالربط بالمخطط أو التحميل إلى المخطط في حقل البيانات الوصفية للمخطط.
- يجب التعبير عن البيانات المجدولة (على سبيل المثال CSV) كمخطط جدول JSON وفقاً للمواصفات المتاحة على الرابط التالي: <http://specs.frictionlessdata.io/json-table-schema/>

يجب أن توفر البيانات المنظمة غير المجدولة (مثل JSON) مخططاً بتنسيق مخطط JSON وفقاً للمواصفات المتاحة على الرابط التالي: <http://json-schema.org/>

مثال: مخطط يمكن قراءته آلياً لملفات CSV

يصف المخطط الصادر بصيغة CSV سمات كل عمود، ويوفر "csvlint.io" واجهة لإجراء ذلك:

الحقول

الفهرس	عنوان العمود	المطلوب	فريد	النوع	قيود القيمة	العنوان/الوصف
1	المعرف	نعم	لا	سلسلة	أقصى طول: 38 أقصى طول: 38 النموذج: \[a-zA-F0-9H8\]{4J-[a-zA-F0-9]{4}-(a-zA-F0-9)(4)-[a-zA-F0-9]{12}\}	كود معاملة فريد
2	السعر	نعم	لا	عدد صحيح غير سالب		سعر مدفوع
3	تاريخ التحويل	نعم	لا	التاريخ الوقت %Y-%m-%d %H:W		بيانات التحويل
4	الكود البريدي	نعم	لا	سلسلة	النموذج: [A-Z]{1,2}[0-9][0-9A-z]?[0-9][A-z]{z}	الكود البريدي للعقار
5	نوع العقار	نعم	لا	سلسلة	النموذج: (D S T F)	نوع العقار (D, S, T, F)
6	قديم/جديد	نعم	لا	سلسلة	النموذج: (Y N)	عقار قديم أم جديد
7	المدة	نعم	لا	سلسلة	النموذج: (F L)	مدة التحويل

وينشئ مخطط يمكن قراءته آلياً بصيغة JSON والتي تبدو كما يلي:

```

{
  "title": "Land Registry Monthly Price Paid Data",
  "description": "Schema for the land registry monthly price-paid data",
  "fields": [
    {
      "name": "ID",
      "description": "Unique transaction code",
      "constraints": {
        "required": true,
        "minLength": 38,
        "maxLength": 38,
        "pattern": "\\{[a-fA-F0-9]{8}-[a-fA-F0-9]{4}-[a-fA-F0-9]{4}-[a-fA-F0-9]{4}-[a-fA-F0-9]{12}\\}"
      }
    },
    {
      "name": "Price",
      "description": "Price Paid",
      "constraints": {
        "required": true,
        "type": "http://www.w3.org/2001/XMLSchema#nonNegativeInteger"
      }
    },
    {
      "name": "Date of Transfer",
      "description": "Data of transfer",
      "constraints": {
        "required": true,
        "type": "http://www.w3.org/2001/XMLSchema#dateTime",
        "datePattern": "%Y-%m-%d %H:%M"
      }
    },
    {
      "name": "Postcode",
      "description": "Postcode for the property",
      "constraints": {
        "required": true,
        "pattern": "[A-Z]{1,2}[0-9][0-9A-Z]? ?[0-9][A-Z]{2}"
      }
    },
    {
      "name": "Property Type",
      "description": "Type of property (D, S, T, F)",
      "constraints": {
        "required": true,
        "pattern": "(D|S|T|F)"
      }
    },
    {
      "name": "Old/New",
      "description": "Old or new property",
      "constraints": {
        "required": true,
        "pattern": "(Y|N)"
      }
    }
  ],
}

```

المصدر: <http://csvlint.io/schemas/530b16c163737676e9260000>

الرموز التعريفية الفريدة

يكون لكل سجل أو صف بيانات رموز تعريفية فريدة لتحسين إمكانية إعادة استخدام مجموعات البيانات. وتتيح الرموز التعريفية الفريدة للمحللين إمكانية دمج مجموعات البيانات بسهولة أكبر.

وفي بعض الحالات، قد تحتاج إلى رموز تعريفية فريدة لبعض السمات أو العناصر داخل المفردات. وهذا يعني أن مجموعات البيانات يمكن دمجها معاً بطرق منطقية - على سبيل المثال، إنشاء سجل لجميع البيانات التي يتم الاحتفاظ بها مقابل عنصر معين "مثل مبنى المدرسة" على مستوى بيانات دبي.

وبالنسبة لأي موضوعات/ أشياء (مثل رموز البلد والمدارس والولايات القضائية والمباني الحكومية) التي يتم الاحتفاظ بها في السجلات الأولية – يجب أن يشير استخدام هذه المواد إلى نفس الرموز التعريفية الفريدة الموجودة في السجل الأساسي ذو الصلة.

المصدر والأولية

هناك حقل بيانات وصفية رئيسي إضافي يتعلق بالمنشأ. ويوثق منشأ البيانات المصادر والمدخلات والمنظمات والأنظمة والعمليات التي شكلت البيانات وأثرت عليها مما يوفر سجلاً تاريخياً للبيانات وأصولها. وهذا بدوره يتيح إمكانية تحليل الاعتماد على البيانات والوعي بالقيود والتغطية وكشف الأخطاء والتدقيق. كما يساعد المستخدمين الآخرين (بما في ذلك الجهة) على فهم القيود ومستوى الثقة التي يمكنهم وضعها في البيانات.

يعتبر تفصيل المصادر والمالكين والأدوات وطرق الجمع والتجهيز إحدى الخطوات الأساسية. ومع مرور الوقت، ينبغي للجهة أن تهدف إلى توحيد وثائق مصدر البيانات وتقديمها بطريقة قابلة للقراءة آلياً (على سبيل المثال باستخدام معيار اتحاد الشبكة العالمية لمصدر البيانات).

وعند نشر البيانات التي تمت معالجتها بشكل كبير، يجب أن تهدف الجهات أيضاً إلى إتاحة بيانات المصدر الأصلية. وهذا يتيح للمستخدمين إمكانية اختيار طرق المعالجة المختلفة لأغراضهم.

الإجراء: يرجى تفصيل المعلومات ذات الصلة حول إنشاء ومعالجة مجموعة البيانات لمساعدة المستخدمين على فهم موثوقية البيانات وتغطيتها وقيودها بشكل أفضل في حقل [البيانات الوصفية](#) "المصدر".

مثال: مكتب الإحصاءات الوطنية في المملكة المتحدة (ONS)

لدى العديد من مجموعات البيانات على الموقع الإلكتروني التابع لمكتب الإحصاءات الوطنية رابط للمنهجية التي تفصل:

- مصدر البيانات (المسح أو الطرف الخارجي وما إلى ذلك)
- حجم العينة (إذا كان عن طريق المسح)
- طريقة جمع البيانات (المقابلات وجها لوجه أو عبر الإنترنت أو طلبات من السلطات)
- الاستثناءات (ما هي البيانات التي لم يتم تضمينها ولماذا)
- طرق التجميع الإحصائي المستخدمة (قمة عدد صغير أو المعدل المتوسط وغيرها)

انظر الأمثلة

<https://www.ons.gov.uk/businessindustryandtrade/internationaltrade/qmis/internationaltradeinservicesqmi>

<https://www.ons.gov.uk/peoplepopulationandcommunity/housing/qmis/housepricestatisticsformallareasqmi>

ملاحظة: حيثما كان ذلك ممكناً، عند نشر مجموعة بيانات مشتقة – وهي مجموعة بيانات ناتجة عن معالجة مجموعات بيانات أخرى، يجب أن يُتاح لمستخدم مجموعة البيانات إمكانية إعادة بنائها من خلال قراءة المنهجية والأجزاء الأصلية في البيانات الوصفية والروابط التالية لبيانات المصدر.

الجدول الزمني

إذا حدث تأخر لمدة سنة أو أشهر بين البيانات التي يتم إنشاؤها وإتاحتها لإعادة استخدامها بحيث يكون لذلك تأثير كبير على قيمة البيانات وإمكانية إعادة استخدامها، يجب على الجهات ضمان إصدار البيانات ونشرها على المنصة في الوقت المناسب – حتى يتم تحديث البيانات وإتاحة أحدث إصدار منها.

يجب على الشخص المسؤول عن جدول نشر كل مجموعة بيانات توثيق التأخير المتوقع الحالي ووضع خطة لكيفية تقليل التأخير بين البيانات التي يتم جمعها وتناولها وإتاحتها على المنصة. ومن الناحية المثالية، لن يكون هناك تأخير أو قد يكون هناك تأخير فقط لعدة أيام، إذا تم الانتهاء من عمليات الفحص والتحضير اللازمة.

الاستيفاء

يجب أن تهدف الجهات إلى إصدار البيانات بأعلى مستوى من التفصيل المتاح وأوسع تغطية (باستثناء الحالات التي ينتهك فيها ذلك الخصوصية أو السرية الشخصية). وهذا بدوره يُوفر للمستخدمين مرونة أكبر واختيار مناسب لتخصيص البيانات بما يلائم احتياجاتهم.

الامتثال

يجب على قائد البيانات التحقق مما يلي:

- أن تحتوي كل مجموعة بيانات على تقرير تقييم جودة بيانات يُنجزه متعهد أو مالك البيانات الذي يورد تفاصيل الجودة الحالية مقابل كل توصية في هذا المعيار وخطة كيفية ونوع التوصيات التي سيتم الوفاء بها في المستقبل وكذلك التواريخ.
- إتمام مراجعة تقرير تقييم الجودة والموافقة عليه من قبل إداري البيانات

الملحق أ:

مصفوفة جودة بيانات دبي

تساعد مصفوفة جودة البيانات الجهات على تقييم مدى موثوقية البيانات وإمكانية الوصول إليها ودقتها وحداثتها وكاملتها والتحقق من صحتها وتوثيقها:
يحدد، لكل من مبادئ جودة البيانات السبعة ، خمسة مستويات من النضج:

- المستوى 1: مبدئي - بيانات غير مُدارة ، بلا مالك ، بلا تنسيق مفتوح ، بلا بيانات وصفية ، إلخ.
- المستوى 2: متوافق جزئياً - مجموعة البيانات لها مالك محدد وتحرز تقدماً نحو التوافق مع معيار جودة البيانات
- المستوى 3: التوافق - مجموعة البيانات تفي بجميع المتطلبات الأساسية المتعلقة بجودة البيانات
- المستوى 4: التحسين - تلبي مجموعة البيانات جميع المتطلبات الأساسية وتقوم أيضاً بتنفيذ ممارسات جيدة إضافية
- المستوى 5: التحسين - جودة البيانات تلبي احتياجات المستخدمين الحاليين والمحتملين في المستقبل ، مع أنظمة واضحة لدفع التحسين المستمر.

المسؤولية والسلطة	1=مبدئي	2=متوافقة جزئياً	3=متوافقة	4=متحسنة	5=متميزة
ليس هناك مسؤول واضح عن مجموعة البيانات يخضع	يتولى مزود البيانات المسؤولية بصفة	مثل المستوى الثاني، بالإضافة إلى:	مثل المستوى الثالث، بالإضافة إلى:	مثل المستوى الرابع، بالإضافة إلى:	

		للمساءلة عنها داخل الجهة. وهناك العديد من مستخدمي البيانات يحتفظون بنسخ مكررة من البيانات ويديرونها بأنفسهم.	شخصية عن جودة البيانات، وقد أجرى تقييمًا مرجعيًا للجودة الحالية للبيانات، مع توثيق مشاكل الجودة المعروفة.	– تواصل مزود البيانات مع المستخدمين الحاليين والمستقبليين المحتملين للبيانات لفهم متطلباتهم بشأن جودة البيانات وتوثيقها ويدير خطة لسد أي فجوات بين مستوى الجودة الحالي والمستوى المطلوب.	– تم وضع آليات لتقديم الملاحظات لتتيح لمستخدمي البيانات طلب تحسين الجودة.	– هناك أدلة واضحة على وجود عمليات فعالة لتمكين التحسين المستمر بناءً على احتياجات المستخدمين.
		لا يمكن للأطراف الأخرى الوصول إلى البيانات بسبب:	ينطبق أحد ما يلي على الأقل على البيانات:	– تم تحديد مجموعات البيانات المرجعية الأساسية التي تستخدمها العديد من الجهات بصورة شائعة من أجل إدراجها كسجل مرجعي ويعمل مزود البيانات مع مؤسسة بيانات دبي لتنفيذ ذلك.	– بالنسبة لمجموعات البيانات التي تستخدمها عدة جهات، تم وضع أنظمة وعمليات لضمان إمكانية توفير البيانات كخدمة لجميع المستخدمين.	– بالنسبة لمجموعات البيانات الأساسية شائعة الاستخدام (وخاصة سجلات دبي)، ليست هناك تكرارات وتتطوي مجموعة البيانات حاليًا على اتفاقيات واضحة لمستوى الخدمة من أجل مستخدمي البيانات.
إمكانية الوصول		– عدم نشرها على الإنترنت أو مشاركتها حاليًا مع جهات أخرى.	– تم نشرها على الإنترنت أو عبر واجهة برمجة تطبيقات متاحة للمستخدمين الخارجيين بصيغة مفتوحة قابلة للقراءة آليًا.	يمكن الوصول إلى البيانات من خلال:	مثل المستوى الثالث، بالإضافة إلى:	مثل المستوى الرابع، بالإضافة إلى ربط مجموعات البيانات بغيرها من البيانات ذات الصلة لتوفير سياق عام لها.
الدقة		المشاكل في دقة مجموعات البيانات (الأخطاء أو الفجوات أو القيود) إما غير معروفة أو من المعلوم أنها كبيرة للغاية.	هناك مشاكل كبيرة في دقة البيانات، لكن تم توثيقها وتوضيحها للمستخدمين.	تم توثيق المشاكل المعروفة في دقة البيانات وتوضيحها للمستخدمين. ويعتبر مستوى الدقة مناسبًا للاستخدام أو الغرض الحالي.	مثل المستوى الثالث، بالإضافة إلى أن الجهة تتواصل بفعالية مع مستخدمي البيانات مدى إمكانية دعم حالات الاستخدام الجديدة من خلال تحسين مستوى الدقة.	مثل المستوى الرابع، بالإضافة إلى أن دقة البيانات مناسبة للغرض المنشود من الاستخدامات الحالية والمحتملة لها، بناءً على الملاحظات الواضحة والموثقة المقدمة من المستخدمين.
الطبيعة الوصفية		لا توجد بيانات وصفية أو مخطط لمجموعة البيانات.	توجد بعض البيانات الوصفية أو مخطط ما لوصف البيانات.	تحتوي مجموعة البيانات على جميع البيانات الوصفية الإلزامية وأي سجلات مرجعية تحتوي على مخطط لها.	تحتوي مجموعة البيانات على جميع البيانات الوصفية الإلزامية ويوجد مخطط لها.	مثل المستوى الرابع، بالإضافة إلى أن مجموعة البيانات تحتوي على جميع البيانات الوصفية الإضافية الموصى بها.
الالتزام بالوقت		مجموعة البيانات قديمة	يتم تحديث مجموعة البيانات بصورة منتظمة لتلبى احتياجات	مثل المستوى الثاني، بالإضافة إلى وجود جدول زمني لنشر مجموعة البيانات يتم الالتزام به كما	مثل المستوى الثالث، بالإضافة إلى وجود ضمانات بتوفير أحدث البيانات في	مثل المستوى الرابع، بالإضافة إلى إدارة تحديثات البيانات بصورة آنية، مع تنفيذ النشر أو المشاركة
بالنسبة لمجموعات						

البيانات المحدثّة	المستخدمين الحاليين.	تمت إضافته إلى البيانات الوصفية لتحديد وتيرة النشر.	المستقبل خلال فترة زمنية محددة.	في نفس الوقت بالنسبة للمستخدمين الداخليين والخارجيين.
بالنسبة لمجموعات البيانات التي تُنشر لمرة واحدة	مجموعة البيانات قديمة، لدرجة عدم وجود فائدة منها.	مجموعة البيانات قديمة، لكن لا تزال لها بعض القيمة لدى المستخدمين.	مجموعة البيانات حديثة بما يكفي لتلبية جميع احتياجات مستخدميها.	[ليس هناك مستوى رابع لمجموعات البيانات التي تُنشر لمرة واحدة]
الاختمال	هناك قدر كبير مفقود من البيانات في مجموعة البيانات أو تعتبر تغطية مجموعة البيانات ضعيفة ولم يتم توثيقها.	هناك بيانات غير موجودة في مجموعة البيانات أو تعتبر تغطيتها ضعيفة، إلا أنه تم توثيقها وتوضيحها لمستخدمي البيانات.	اكتمال البيانات مناسب للأغراض الحالية، وتعتبر مجموعة البيانات مفهومة، ويمكن استخدامها بمفردها أو مع البيانات المرجعية، ويتم توثيق أي سجلات مفقودة أو فجوات وتوضيحها.	مثل المستوى الثالث، بالإضافة إلى أن الجهة تتواصل بفعالية مع مستخدمي البيانات المحتملين لفهم مدى إمكانية دعم حالات الاستخدام الجديدة من خلال تحسين تغطية البيانات.
التحقق من الصحة	لا يتم التحقق من صحة البيانات	تم استخدام المفردات أو الخانات المعتمدة عند إنشاء / تسجيل البيانات (على سبيل المثال، ضمان استخدام صيغة مطابقة متفق عليها لأرقام الهواتف).	مثل المستوى الثاني، بالإضافة إلى التحقق حالياً من جميع الخانات التي لا تتطلب نصاً حراً (على سبيل المثال، البحث عن العنوان باستخدام الرمز البريدي، والتحقق من إدخال رقم في خانات الأرقام).	مثل المستوى الرابع، بالإضافة إلى مراجعة البيانات بصفة منتظمة لحذف السجلات المكررة وعلاج الأخطاء في أنظمة البيانات بالكامل.
		وتلتزم الجهة باصطلاحات التسمية وصيغة السجل المرجعي بالنسبة للبيانات المرجعية شائعة الاستخدام.	مثل المستوى الثالث، بالإضافة إلى أن مجموعة البيانات تستخدم سجلات دبي المناسبة وتتحقق من البيانات وفقاً لها. ويتم ترميز التحقق من صحة البيانات في مخطط يمكن خلاله التحقق من صحة البيانات تلقائياً.	