



Preuve de concept de restitution automatisée (ISO-19157)

CNIG - Réunion QuaDoGéo - 19 Mai 2021
Mathieu Rajerison - Cerema Med

Le POC précédent



Preuve de concept de restitution automatisée (FR)

CNIG - Réunion QuaDoGéo - 2 Octobre 2020
Mathieu Rajerison - Cerema Med

https://docs.google.com/presentation/d/1TCYm14_mcmfzSNTyCeLvuT42KlrhgTr3vMO6HzNbLOg/edit

2.1 Exhaustivité

2.1.1 Taux d'exhaustivité

Le taux d'exhaustivité est la mesure de la qualité des données portant sur l'exhaustivité retenue par les fiches méthodologiques du GT CNIG QuaDoGéo.

Il s'agit d'une mesure qui n'est pas décrite dans la norme ISO 19157.

Ligne	Composant	Description
1	Nom	Taux d'exhaustivité
2	Alias	txEx
3	Critère / sous-critère	Exhaustivité / Taux d'exhaustivité
4	Mesure de base	taux d'erreur
5	Définition	Nombre total d'éléments en excès ou manquant dans le jeu de données par rapport au nombre total d'objets du jeu de données.
6	Description	Soient : N_n : nombre d'objets constaté dans l'échantillon ou dans le jeu de données Nb_E : nombre d'objets en excédent dans l'échantillon ou dans le jeu de données Nb_O : nombre d'objets manquant dans l'échantillon ou dans le jeu de données $N_i = N_n + Nb_O - Nb_E$: nombre d'objets attendus dans l'échantillon ou dans le jeu de données Alors, le taux d'exhaustivité vaut : $1 - \frac{(Nb_O + Nb_E)}{N_i}$
7	Paramètre	-
8	Type de valeur	Nombre réel, pourcentage
9	Structure de la valeur	-
10	Référence de la source	Fiches méthodologique « Critère d'exhaustivité » Non référencé dans ISO 19157
11	Exemple	Le jeu de données comprend 50 objets, 5 sont en excédents, 2 sont manquants. Le jeu de données devrait donc théoriquement contenir $50+2-5 = 47$ objets. Le taux d'exhaustivité est alors égal à $1 - (7/47)$ soit : 85 %
12	Identifiant	https://data.geocatalogue.fr/mc/qualiteMesures/txEx

QuaDoGeo Report

Dataset

My Dataset

Dataset quality metrics

exhaustivity : ★★★★★★★★★★

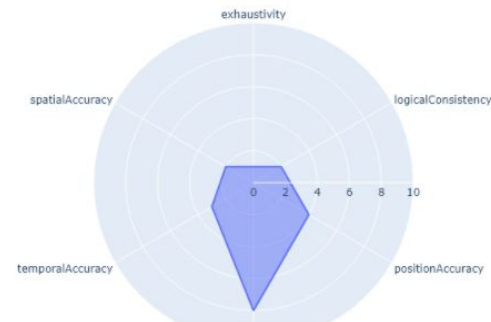
logicalConsistency : ★★★★★★★★★★

positionAccuracy : ★★★★★★★★★★

thematicAccuracy : ★★★★★★★★★★

temporalAccuracy : ★★★★★★★★★★

spatialAccuracy : ★★★★★★★★★★



Les ressources

[Norme ISO-19157](#) (github)

[Registre français des méta-données](#) (CNIG)

[Implémentation Géocatalogue](#) (Géocatalogue)

Le POC précédent

```
C:\python\python quadorender.py report.xml
```

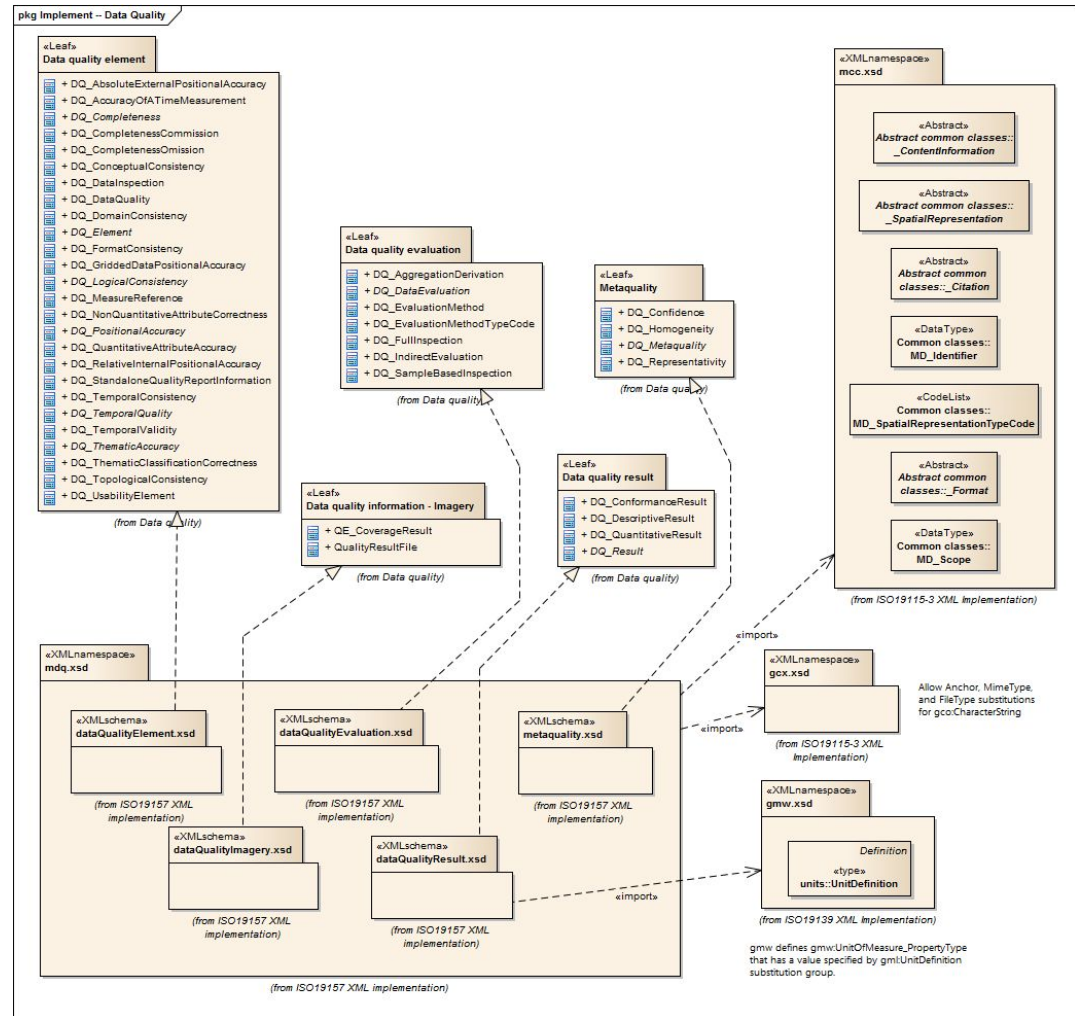
Install' et utilisation

<https://gitlab.cerema.fr/mathieu.rajerison/quadogeo/>

ISO 19157

Metadata for Data Quality

<https://github.com/ISO-TC211/XML/blob/master/standards.iso.org.annotated/iso/19157/-2/mdq/1.0/index.html>



ISO 19157

Metadata for Data Quality

<https://github.com/ISO-TC211/XML/blob/master/standards.iso.org.annotated/iso/19157/-2/mdq/1.0/index.html>

pkg Implement -- Data Quality

«Leaf»

Data quality element

- + DQ_AbsoluteExternalPositionalAccuracy
- + DQ_AccuracyOfATimeMeasurement
- + DQ_Completeness
- + DQ_CompletenessCommission
- + DQ_CompletenessOmission
- + DQ_ConceptualConsistency
- + DQ_DataInspection
- + DQ_DataQuality
- + DQ_DomainConsistency
- + DQ_Element
- + DQ_FormatConsistency
- + DQ_GriddedDataPositionalAccuracy
- + DQ_LogicalConsistency
- + DQ_MeasureReference
- + DQ_NonQuantitativeAttributeCorrectness
- + DQ_PositionalAccuracy
- + DQ_QuantitativeAttributeAccuracy
- + DQ_RelativeInternalPositionalAccuracy
- + DQ_StandaloneQualityReportInformation
- + DQ_TemporalConsistency
- + DQ_TemporalQuality
- + DQ_TemporalValidity
- + DQ_ThematicAccuracy
- + DQ_ThematicClassificationCorrectness
- + DQ_TopologicalConsistency
- + DQ_UsabilityElement

(from Data Quality)

«l
D
E

Les mesures

DQ_AbsoluteExternalPositionalAccuracy

DQ_AccuracyOfATimeMeasurement

AbstractDQ_Completeness

DQ_CompletenessCommission

DQ_CompletenessOmission

DQ_ConceptualConsistency

DQ_DataInspection

DQ_DataQuality

DQ_DomainConsistency

AbstractDQ_Element

DQ_FormatConsistency

DQ_GriddedDataPositionalAccuracy

AbstractDQ_LogicalConsistency

DQ_MeasureReference

DQ_NonQuantitativeAttributeCorrectness

AbstractDQ_PositionalAccuracy

DQ_QuantitativeAttributeAccuracy

DQ_RelativeInternalPositionalAccuracy

DQ_StandaloneQualityReportInformation

DQ_TemporalConsistency

AbstractDQ_TemporalQuality

DQ_TemporalValidity

AbstractDQ_ThematicAccuracy

DQ_ThematicClassificationCorrectness

DQ_TopologicalConsistency

DQ_UsabilityElement

La norme ISO 19157

```
<!-- Schema document created by ShapeChange - http://shapechange.net -->
<element name="DQ_AbsoluteExternalPositionalAccuracy" substitutionGroup="mdq:AbstractDQ_F
...<annotation>
...<documentation>closeness of reported coordinate values to values accepted as or being
...</annotation>
</element>
<complexType name="DQ_AbsoluteExternalPositionalAccuracy_Type">
...<comple
...<exte
...<se
...</extension>
...</complexContent>
</complexType>
<complexType name="DQ_AbsoluteExternalPositionalAccuracy_PropertyType">
...<sequence minOccurs="0">
...<element ref="mdq:DQ_AbsoluteExternalPositionalAccuracy"/>
...</sequence>
...<attributeGroup ref="gco:ObjectReference"/>
...<attribute ref="gco:nilReason"/>
```

[dataQualityElements.xsd](#)

report-19157.xml

```
<?xml version="1.0"?>...  
<quadogeo>  
  <dataset>My·Dataset</dataset>  
  <metrics>  
    <!--Exhaustivy,·completeness-->  
    <AbstractDQ_Completeness·group="exhaustivity"·valueType="Real">0.6</AbstractDQ_Completeness>  
    <DQ_CompletenessCommission·group="exhaustivity"·valueType="Real">0.6</DQ_CompletenessCommission>  
    <!--Logical-->  
    <DQ_ConceptualConsistency·group="logicalConsistency"·valueType="Real">0.2</DQ_ConceptualConsistency>  
    <DQ_DataQuality·group="logicalConsistency"·valueType="Real">0.2</DQ_DataQuality>  
    <DQ_FormatConsistency·group="logicalConsistency"·valueType="Real">0.2</DQ_FormatConsistency>  
    <AbstractDQ_LogicalConsistency·group="logicalConsistency"·valueType="Real">0.2</AbstractDQ_LogicalConsistency>  
    <DQ_NonQuantitativeAttributeCorrectness·group="logicalConsistency"·valueType="Real">0.2</DQ_NonQuantitativeAttributeCorrectness>  
    <DQ_TopologicalConsistency·group="thematicAccuracy"·valueType="Real">0.2</DQ_TopologicalConsistency>  
    <!--PositionalAccuracy-->  
    <DQ_AbsoluteExternalPositionalAccuracy·group="positionalAccuracy"·valueType="Real">0.2</DQ_AbsoluteExternalPositionalAccuracy>  
    <AbstractDQ_PositionalAccuracy·group="positionalAccuracy"·valueType="Real">0.2</AbstractDQ_PositionalAccuracy>  
    <DQ_RelativeInternalPositionalAccuracy·group="positionalAccuracy"·valueType="Real">0.2</DQ_RelativeInternalPositionalAccuracy>  
    <!--ThematicAccuracy-->  
    <DQ_DomainConsistency·group="thematicAccuracy"·valueType="Real">0.2</DQ_DomainConsistency>  
    <DQ_QuantitativeAttributeAccuracy·group="thematicAccuracy"·valueType="Real">0.2</DQ_QuantitativeAttributeAccuracy>  
    <AbstractDQ_ThematicAccuracy·group="thematicAccuracy"·valueType="Real">0.6</AbstractDQ_ThematicAccuracy>  
    <!--TemporalAccuracy-->  
    <DQ_TemporalConsistency·group="temporalAccuracy"·valueType="Real">0.2</DQ_TemporalConsistency>  
    <AbstractDQ_TemporalConsistency·group="temporalAccuracy"·valueType="Real">0.2</AbstractDQ_TemporalConsistency>  
    <DQ_TemporalValidity·group="temporalAccuracy"·valueType="Real">0.2</DQ_TemporalValidity>  
    <DQ_AccuracyOfATimeMeasurement·group="temporalAccuracy"·valueType="Real">0.2</DQ_AccuracyOfATimeMeasurement>  
    <!--SpatialAccuracy-->  
    <DQ_GriddedDataPositionalAccuracy·group="spatialAccuracy"·valueType="Real">0.8</DQ_GriddedDataPositionalAccuracy>  
    <!--Usability-->  
    <DQ_UsabilityElement·group="usability"·valueType="Integer">1</DQ_UsabilityElement>
```

```
<?xml version="1.0"?>...  
<quadogeo>  
  <dataset>My·Dataset</dataset>  
  <metrics>  
    <!--Exhaustivy,·completeness-->  
    <AbstractDQ_Completeness·group="exhaustivity"·valueType="Real">0.6</AbstractDQ_Completeness>  
    <DQ_CompletenessCommission·group="exhaustivity"·valueType="Real">0.6</DQ_CompletenessCommission>  
    <!--Logical-->  
    <DQ_ConceptualConsistency·group="logicalConsistency"·valueType="Real">0.2</DQ_ConceptualConsistency>  
    <DQ_DataQuality·group="logicalConsistency"·valueType="Real">0.2</DQ_DataQuality>  
    <DQ_FormatConsistency·group="logicalConsistency"·valueType="Real">0.2</DQ_FormatConsistency>  
    <AbstractDQ_LogicalConsistency·group="logicalConsistency"·valueType="Real">0.2</AbstractDQ_LogicalConsistency>
```

report-19157.xml

Le script

```
C:\python\python quadorender.py report-1957.xml
```

➔ report.html

Dataset

My Dataset

Date de création : 2020-09-29

Date de révision : 2020-09-29

Date de publication : 2020-09-29

User Advice

note ★★★★★

Details

exhaustivity ★★★★★

logicalConsistency ★★★★★

positionAccuracy ★★★★★

thematicAccuracy ★★★★★

temporalAccuracy ★★★★★

spatialAccuracy ★★★★★

usability ★★★★★

Dataset quality metrics

exhaustivity ★★★★★

logicalConsistency ★★★★★

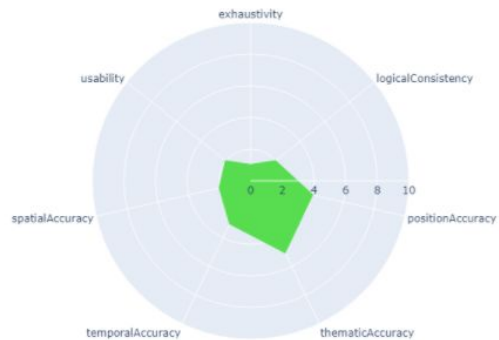
positionAccuracy ★★★★★

thematicAccuracy ★★★★★

temporalAccuracy ★★★★★

spatialAccuracy ★★★★★

usability ★★★★★



Metrics depending on usage

Spatial statistics and reports

exhaustivity ★★★★★
logicalConsistency ★★★★★
positionAccuracy ★★★★★
thematicAccuracy ★★★★★
temporalAccuracy ★★★★★★
spatialAccuracy ★★★★★★★
usability ★★★★★



Emergency and rescue

exhaustivity ★★★★★
logicalConsistency ★★★★★

Mesures

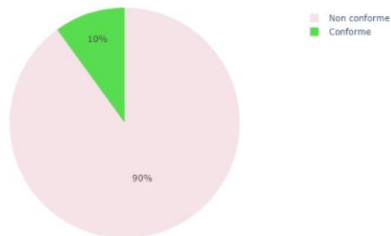
logicalConsistency

DQ_ConceptualConsistency True

DQ_DataQuality Lorem

DQ_FormatConsistency 0.1

DQ_FormatConsistency



AbstractDQ_LogicalConsistency False

DQ_NonQuantitativeAttributeCorrectness True

positionAccuracy

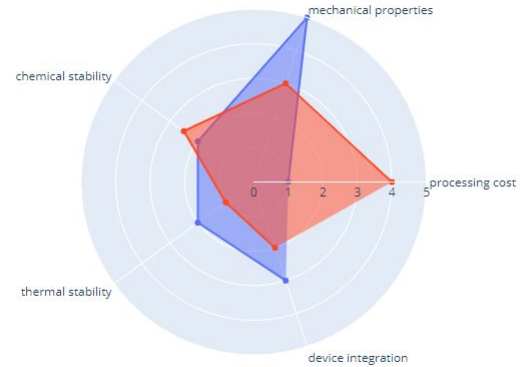
DQ_AbsoluteExternalPositionalAccuracy 0.2

DQ_AbsoluteExternalPositionalAccuracy



Evolutions possibles

- Radars empilés pour rendre les notes calculées vs les notes selon les usages



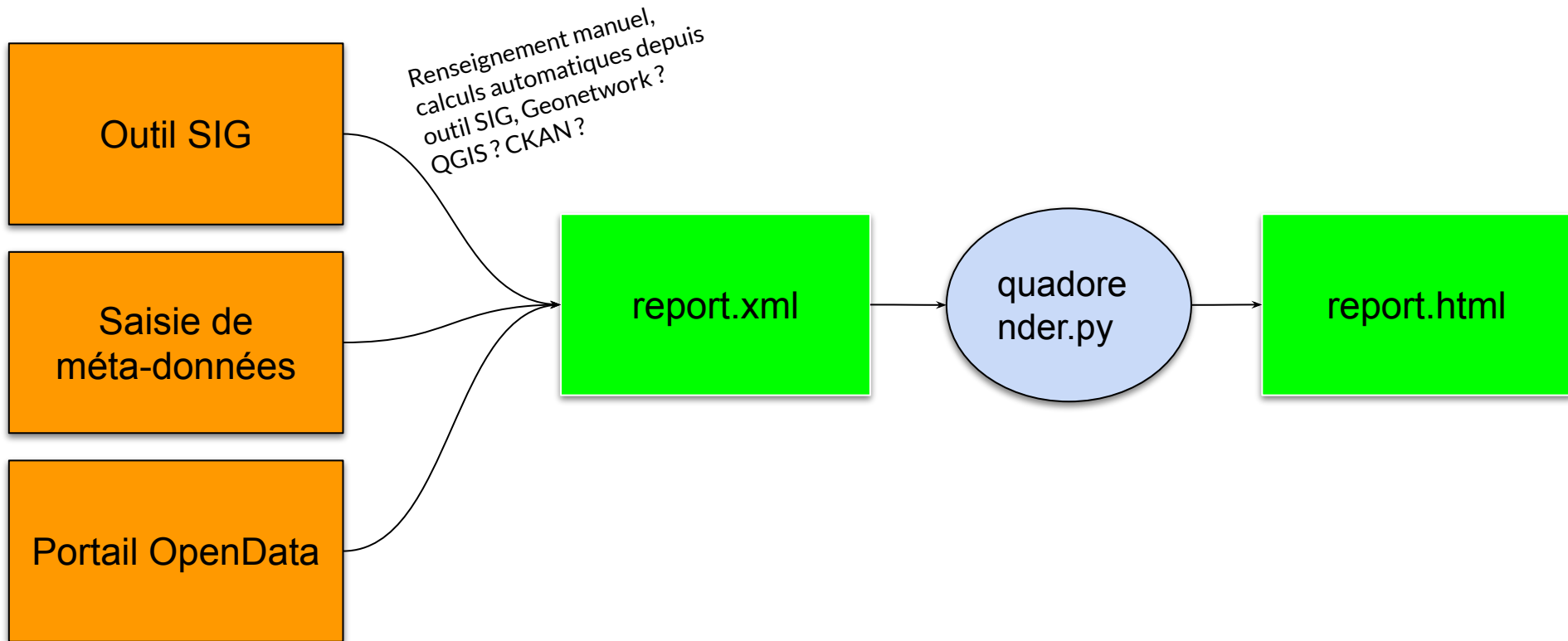
? Méthodo

- Comment **attribuer la note synthétique** en fonction des métriques : grille de calcul ?
- Comment **rendre** les nombres : la relativité est importante. La représentation absolue est-elle utile ?
- Un référentiel des **usages** ?
- **Lien entre le data.geocatalogue et la norme 19157**, notamment le nom des indicateurs

? XML ISO-19157

- Quel est le type de chaque métrique ISO-19157 ? ([Lien avec dqm.xml ?](#))
- Existe-t-il des exemples de rendus XML ISO-19157 ?

Les briques



Conclusion

Contribuer :

<https://gitlab.cerema.fr/mathieu.rajerison/quadogeo/>

Test d'implémentation dans un portail OpenData ?