



Preuve de concept de restitution automatisée

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

Le POC

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

Les inputs

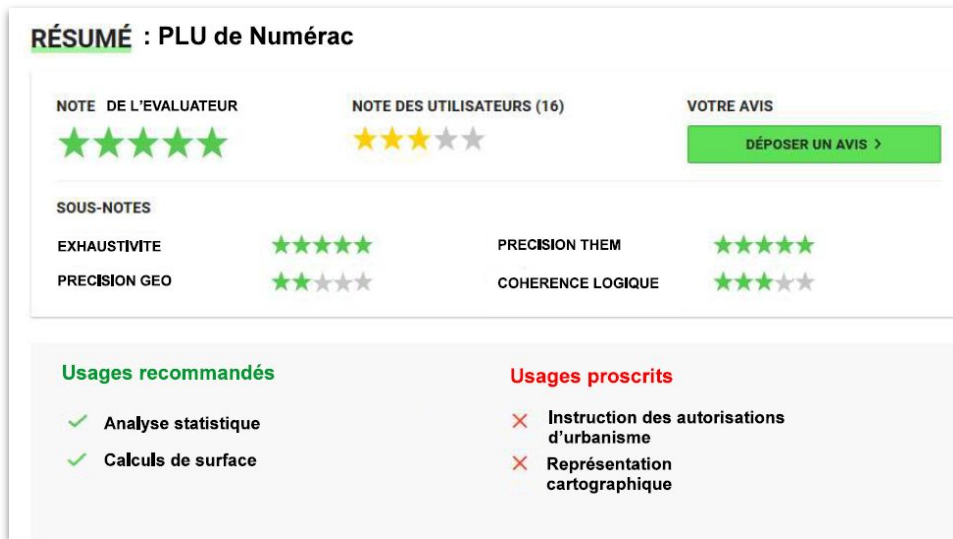
Les méta-données



The image shows the cover of a report titled 'Métadonnées relatives à la qualité des données géographiques' (Metadata related to the quality of geographical data). It is part of the 'REGISTRE FRANÇAIS' (French Register) and is labeled as a 'Version Projet' (Project Version) from October 2019. The report is published by the 'CNIG' (National Institute for Geographic Information and Forestry) and is associated with 'Cerema'.

Logo of the French Republic and CNIG (Centre National de l'Information Géographique) are visible in the top left. The Cerema logo is in the top right. The title 'Métadonnées relatives à la qualité des données géographiques' is centered. Below it, 'REGISTRE FRANÇAIS' is written in bold. At the bottom, 'Version Projet' is highlighted in yellow, followed by 'Registre CNIG' and 'octobre 2019'.

Maquette



The image is a screenshot of a web interface for the 'PLU de Numérac' project. It features a 'RÉSUMÉ' (Summary) section with three main components: 'NOTE DE L'EVALUATEUR' (Evaluator's Note) showing 5 green stars, 'NOTE DES UTILISATEURS (16)' (User Note) showing 4 yellow stars and 2 grey stars, and 'VOTRE AVIS' (Your Review) with a green button labeled 'DÉPOSER UN AVIS >'. Below this is a 'SOUS-NOTES' (Sub-Notes) section with four categories: 'EXHAUSTIVITE' (5 green stars), 'PRECISION THEM' (5 green stars), 'PRECISION GEO' (3 green stars, 2 grey stars), and 'COHERENCE LOGIQUE' (3 green stars, 2 grey stars). At the bottom, there are two columns: 'Usages recommandés' (Recommended uses) with a checkmark for 'Analyse statistique' and 'Calculs de surface', and 'Usages proscrits' (Prohibited uses) with a red X for 'Instruction des autorisations d'urbanisme' and 'Représentation cartographique'.

RÉSUMÉ : PLU de Numérac

NOTE DE L'EVALUATEUR: ★★★★★

NOTE DES UTILISATEURS (16): ★★☆☆☆

VOTRE AVIS: DÉPOSER UN AVIS >

SOUS-NOTES

EXHAUSTIVITE	★★★★★	PRECISION THEM	★★★★★
PRECISION GEO	★★★☆☆	COHERENCE LOGIQUE	★★★☆☆

Usages recommandés

- ✓ Analyse statistique
- ✓ Calculs de surface

Usages proscrits

- ✗ Instruction des autorisations d'urbanisme
- ✗ Représentation cartographique

Les méta-données

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_m : 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 $N_i = N_m + Nb_O - Nb_E$: 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/ncl/qualiteMesures/txEx

Proposition de XML

- Des **métriques** : les indicateurs calculés
- Des **notes** de qualité calculées depuis les métriques
- Des **notes** de qualité selon les **usages**

Proposition de XML

```
<?xml version="1.0" encoding="UTF-8" ?>
<quadgeo>
  <dataset>My Dataset</dataset>
  <metrics>
    <txEx group="exhaustivity" valueType="Real">0.1</txEx>
    <NonConfConceptuelle group="logicalConsistency" valueType="Boolean">False</NonConfConceptuelle>
    <txConfDomVal group="logicalConsistency" valueType="Real">0.2</txConfDomVal>
    <confStructPhysiq group="logicalConsistency" valueType="Boolean">False</confStructPhysiq>
    <tauxErrFormat group="logicalConsistency" valueType="Real">0.1</tauxErrFormat>
    <tauxConnexErr group="logicalConsistency" valueType="Real">0.5</tauxConnexErr>
    <nbMicroSurfErr group="logicalConsistency" valueType="Integer">10</nbMicroSurfErr>
    <nbBoucle group="logicalConsistency" valueType="Integer">20</nbBoucle>
    <nbChevauchErr group="logicalConsistency" valueType="Integer">30</nbChevauchErr>
    <ClassePrecisionArrrete2003 group="positionAccuracy" valueType="Text">Classe 1</ClassePrecisionArrrete2003>
    <valMoyIncertitude group="positionAccuracy" valueType="Real">0.3</valMoyIncertitude>
    <errCirculaireP95 group="positionAccuracy" valueType="Real">0.5</errCirculaireP95>
    <ErrVerticaleRelative group="positionAccuracy" valueType="Real">0.4</ErrVerticaleRelative>
    <ErrHorizontaleRelative group="positionAccuracy" valueType="Real">0.3</ErrHorizontaleRelative>
    <RMCM group="thematicAccuracy" valueType="Matrix">50,30,20</RMCM>
    <kappa group="thematicAccuracy" valueType="Integer">-1</kappa>
    <tauxValAttOk group="thematicAccuracy" valueType="Real">0.3</tauxValAttOk>
    <incertitudeValAtt595 group="thematicAccuracy" valueType="Integer">50</incertitudeValAtt595>
    <exactMesTemps095 group="temporalAccuracy" valueType="Integer">50</exactMesTemps095>
    <tauxConfChrono group="temporalAccuracy" valueType="Real">0.3</tauxConfChrono>
    <txConfDomVal group="temporalAccuracy" valueType="Real">0.2</txConfDomVal>
    <genealogie group="genealogy" valueType="Text">lorem</genealogie>
    <resolutionSpatiale group="spatialAccuracy" valueType="Text">lorem</resolutionSpatiale>
    <etendueTemporelle group="temporalAccuracy" valueType="Date">2020-09-29</etendueTemporelle>
    <dateCreation group="referenceDates" valueType="Date">2020-09-29</dateCreation>
    <dateRevision group="referenceDates" valueType="Date">2020-09-29</dateRevision>
    <datePublication group="referenceDates" valueType="Date">2020-09-29</datePublication>
  </metrics>
  <synthese>
    <exhaustivity>1</exhaustivity>
    <logicalConsistency>2</logicalConsistency>
    <positionAccuracy>4</positionAccuracy>
    <thematicAccuracy>8</thematicAccuracy>
    <temporalAccuracy>3</temporalAccuracy>
    <spatialAccuracy>2</spatialAccuracy>
  </synthese>
  <usages>
    <usage type="survey" description="Spatial statistics and reports">
      <exhaustivity>4</exhaustivity>
      <logicalConsistency>3</logicalConsistency>
      <positionAccuracy>2</positionAccuracy>
      <thematicAccuracy>1</thematicAccuracy>
      <temporalAccuracy>7</temporalAccuracy>
      <spatialAccuracy>9</spatialAccuracy>
    </usage>
    <usage type="emergency" description="Emergency and rescue">
      <exhaustivity>2</exhaustivity>
      <logicalConsistency>2</logicalConsistency>
      <positionAccuracy>8</positionAccuracy>
      <thematicAccuracy>4</thematicAccuracy>
      <temporalAccuracy>3</temporalAccuracy>
      <spatialAccuracy>2</spatialAccuracy>
    </usage>
  </usages>
</quadgeo>
```

```
<?xml version="1.0" encoding="UTF-8" ?>
<quadgeo>
  <dataset>My Dataset</dataset>
  <metrics>
    <txEx group="exhaustivity" valueType="Real">0.1</txEx>
    <NonConfConceptuelle group="logicalConsistency" valueType="Boolean">False</NonConfConceptuelle>
    <txConfDomVal group="logicalConsistency" valueType="Real">0.2</txConfDomVal>
    <confStructPhysiq group="logicalConsistency" valueType="Boolean">False</confStructPhysiq>
```

```
</metrics>
  <synthese>
    <exhaustivity>1</exhaustivity>
    <logicalConsistency>2</logicalConsistency>
    <positionAccuracy>4</positionAccuracy>
    <thematicAccuracy>8</thematicAccuracy>
    <temporalAccuracy>3</temporalAccuracy>
    <spatialAccuracy>2</spatialAccuracy>
  </synthese>
</quadgeo>
```

```
</usage>
  <usage type="emergency" description="Emergency and rescue">
    <exhaustivity>2</exhaustivity>
    <logicalConsistency>2</logicalConsistency>
    <positionAccuracy>8</positionAccuracy>
    <thematicAccuracy>4</thematicAccuracy>
    <temporalAccuracy>3</temporalAccuracy>
    <spatialAccuracy>2</spatialAccuracy>
  </usage>
</usages>
```

Types de valeurs et rendus

- Des **métriques** : les indicateurs calculés
 - Taux > pie charts
 - Nombres > texte ou bar charts
 - Texte > texte
 - Booléens > Rouge ou Vert
- Les **notes**
 - Etoiles

Librairies utilisées

- Langage : python
- Librairies :
 - plotly (graphiques interactifs)
 - mais aussi pandas
 - et des dépendances : psutil, orca, kaleido

Le script

- quadorender.py
- Exécution :

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

- Les sorties
 - report.html
 - plots au format PNG (intégrés au report.html)

La sortie : report.html

QuaDoGeo Report

Dataset

My Dataset

Dataset quality metrics

exhaustivity : ★★★★★★★★★★

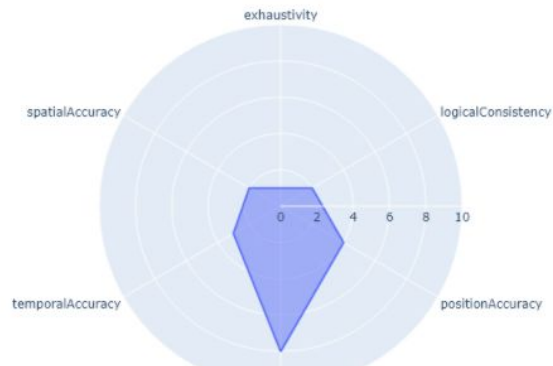
logicalConsistency : ★★★★★★★★★★

positionAccuracy : ★★★★★★★★★★

thematicAccuracy : ★★★★★★★★★★

temporalAccuracy : ★★★★★★★★★★

spatialAccuracy : ★★★★★★★★★★



La sortie : report.html

Metrics depending on usage

Spatial statistics and reports

exhaustivity : ★★★★★★★★

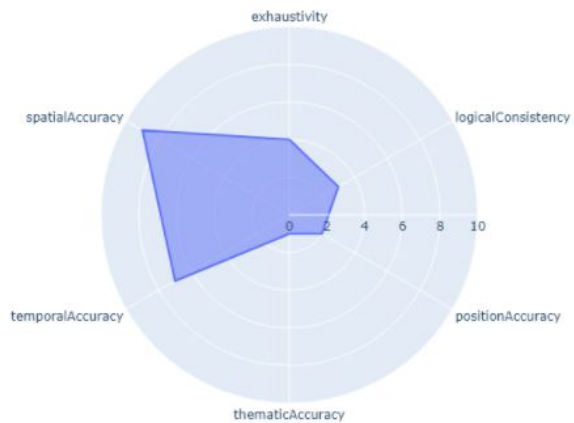
logicalConsistency : ★★★★★★★★

positionAccuracy : ★★★★★★★★

thematicAccuracy : ★★★★★★★★

temporalAccuracy : ★★★★★★★★

spatialAccuracy : ★★★★★★★★

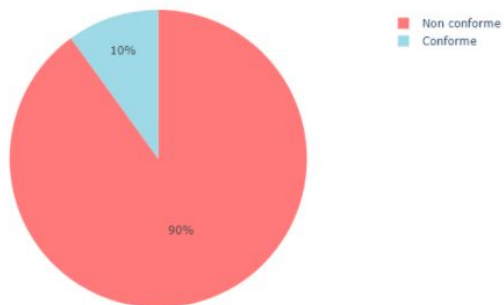


La sortie : report.html

exhaustivity

txEx : 0.1

txEx



logicalConsistency

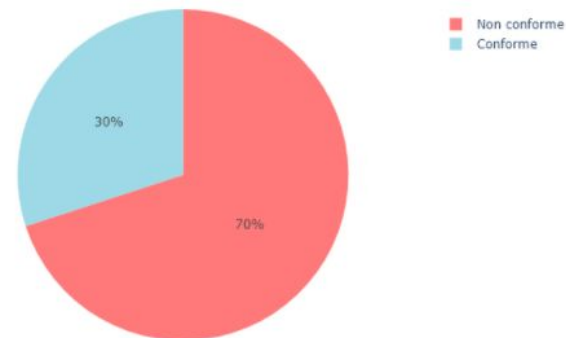
NonConfConceptuelle : False

positionAccuracy

ClassePrecisionArrete2003 : Classe 1

valMoyIncertitude : 0.3

valMoyIncertitude



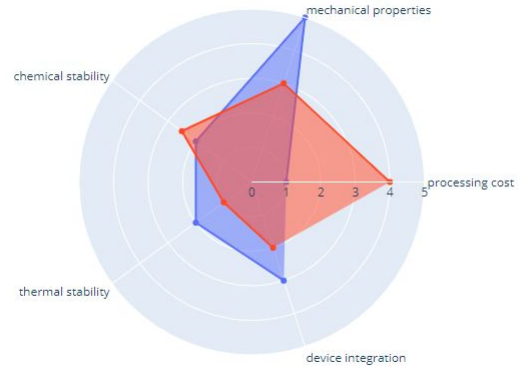
Le script : éléments techniques

- Des fonctions utiles :
 - `render_plot_real` > pie charts
 - `render_radar` > radar charts
 - `render_stars` > étoiles

```
8
9 # Render stars
0 def render_stars(nStars) :
1     nNonStars = 10 - nStars
2     elt_checked = '<span class="fa fa-star checked"></span>'
3     elt_notChecked = '<span class="fa fa-star"></span>'
4     elts_checked = '\n'.join([elt_checked]*nStars)
5     elts_notChecked = '\n'.join([elt_notChecked]*nNonStars)
6     elts = "%s%s"%(elts_checked, elts_notChecked)
7     return(elts)
8
```

Evolutions possibles

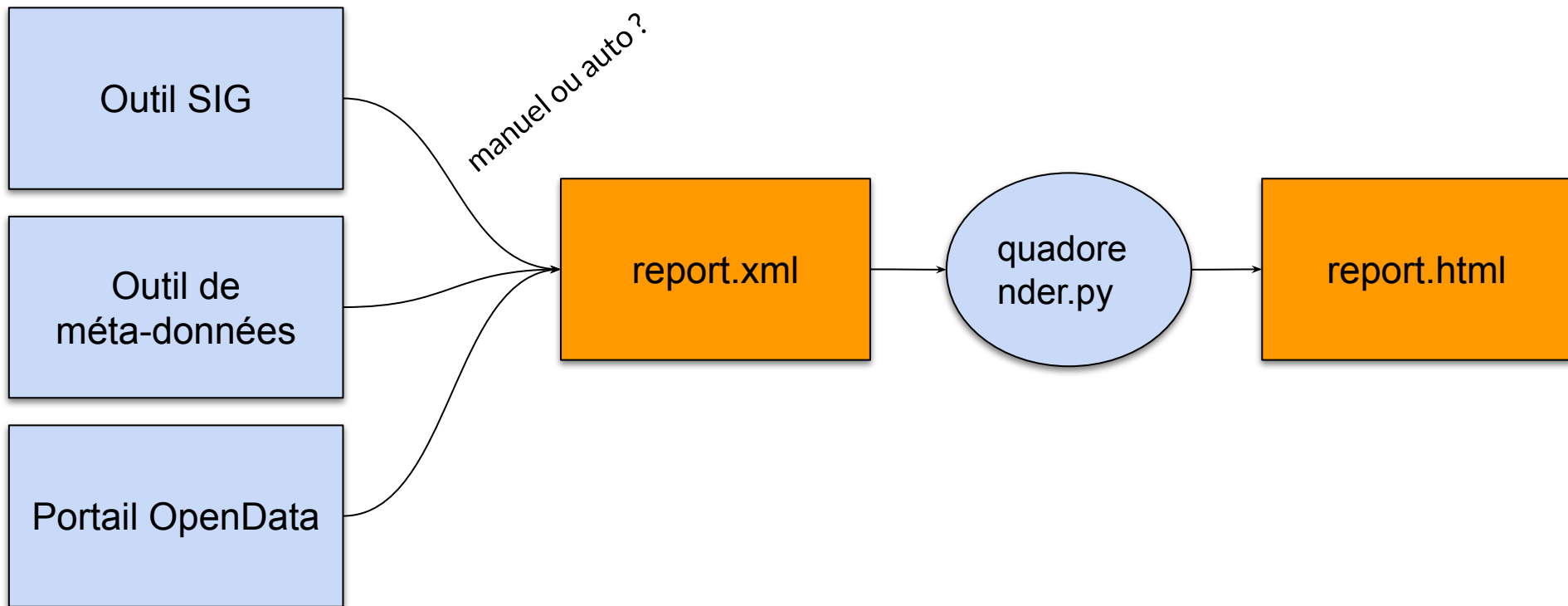
- *Juste un POC* : feuilles de **style** pour un rendu plus joli
- Radars **empilés** pour rendre les notes calculées vs les notes selon les usages



Interrogations

- Comment **noter** en fonction des métriques : grille de calcul ?
- Comment **rendre** les nombres : la relativité est importante
- Un référentiel des **usages**
- Implémentation : Intégration dans un framework : grandlyon ?
- XML : Comment produire le XML :
 - renseignement manuel, calculs automatiques depuis outil SIG ?
 - depuis geonetwork ? QGIS ? CKAN ?

Les briques



Conclusion

Contribuer aux outils :

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

Test d'implémentation dans un portail OpenData ?