



Cerema

Concepts et indicateurs de la qualité des données géographiques

Du bon usage de la
norme 19157

Mathieu Rajerison
CEREMA Med > DCEDI > AGIL



Cerema Méditerranée

Mardi 6 Février 2018- Séminaire « Outils et organisations pour la qualification des données géographiques »

A quoi servent les normes ?

- La norme sert de **base** dans les relations entre **partenaires**.
- Référencée dans un contrat, elle **s'impose** aux parties.

Qu'est ce que l'ISO ?

- ISO, fédération mondiale d'organismes nationaux de normalisation.
- L'élaboration des normes est confiée à un comité technique.



Les enjeux

- Des données de plus en plus **partagées** et **échangées**.
- Nécessité de **comparer** des jeux de données et sélectionner **le plus adapté** aux **exigences** de son application.
- **Vérifier la réponse** aux critères définis dans la spécification du produit.
- Avoir **une compréhension commune** des mesures de qualité

Et la norme 19157 ?

- Adoptée le 9 Novembre 2013, il s'agit d'une norme **européenne** pour l'évaluation de la qualité des données géographiques. Elle agrège les normes ISO 19113, 19114, 19138.
- Elle **normalise** les composants et structures des mesures de la qualité des données et fournit les lignes directrices pour les procédures d'évaluation **des informations de qualité quantitatives** relatives aux données géographiques.

Composition de la commission de normalisation

Des agents d'organismes publics

Des universitaires

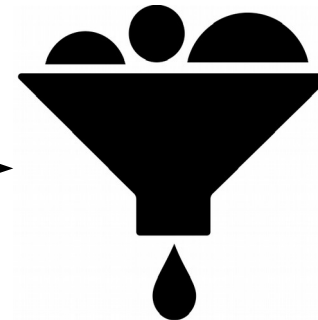
Des membres d'associations

Des agents de sociétés privées



Mathieu Rajerison, Cerema Méditerranée
Concepts et indicateurs de la qualité des données géographiques, du bon usage de la norme 19157
Mardi 6 Février 2018 – Séminaire « Outils et organisations pour la qualification des données géographiques »

[10]	Règles de schéma d'application
[11]	Méthodologie de catalogage des entités
[12]	Schéma de la géométrie et des fonctions de couverture
[13]	Spécifications de contenu informationnel
[13]	Information — Vocabulaire — Partie 4: Organisation des données
[14]	Information — Registres de métadonnées (RM) — Partie 3: Métamodèle de registre et
[15]	Information — Structure de données pour les images, les matrices et les mosaïques
[16]	Information — Métadonnées — Implémentation de schémas XML
[17]	ISO 19156:2011, Information géographique — Observations et mesures
[18]	CRC Handbook of Tables for Probability and Statistics, 2nd edition, 1982 (Guide CRC des tables de probabilités et de statistiques)
[19]	Department of Defense (US). (Ministère américain de la Défense) Standard Practice: Mapping, Charting and Geodesy, Accuracy (Pratique normalisée: Cartographie terrestre et marine, géodésie, Exactitude MIL STD 600001, 1990)
[20]	NATO STANAG 2215 IGEO, Evaluation of land maps, aeronautical charts and digital topographic data, 6th edition (OTAN — STANAG — 2215 IGEO — Évaluation des cartes terrestres, des cartes aéronautiques et des données topographiques numérisées, 6e édition)
[21]	WIKIPEDIA http://en.wikipedia.org/wiki/Hypergeometric_distribution (loi hypergéométrique)



Un contenu :

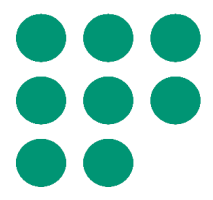
- extrêmement riche (160 pages)
- basé sur des sources avérées (normes ISO déjà existantes)
- pointu (documents issus de l'armée)
- conçu par des experts reconnus dans le domaine.

...qui peut aussi sembler austère.

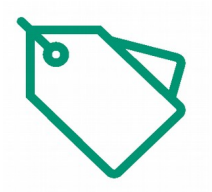
Ayant connaissance des principaux besoins exprimés dans les services, on pourrait **extraire** de la norme ce qui semble le plus utile, éventuellement le **synthétiser** et le **présenter de façon plus attrayante** (utilisation d'illustrations, d'exemples...)

Les critères de Qualité des données géographiques s'intéressent aussi bien aux dimensions géographique (positionnement), géométrique (topologie,...) qu'aux attributs des objets.

Exhaustivité

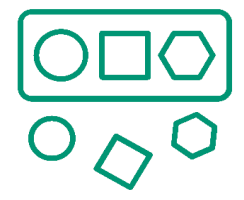


Précision thématique



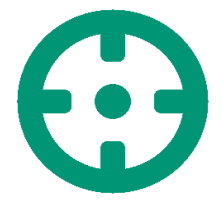
Attributs

Cohérence logique



Géométrie
Attributs

Précision de position

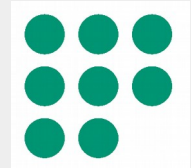


Géométrie

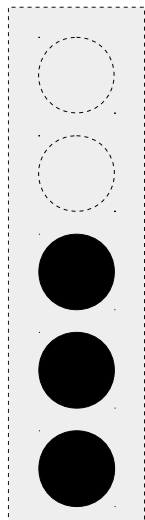
Cohérence temporelle



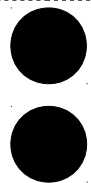
Attributs



Ce critère rassemble tout ce qui caractérise l'absence de données attendues ou leur présence à tort.



Éléments
manquants
NbE



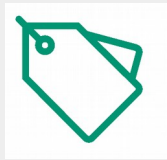
Éléments
en excès
NbO

$$\text{taux d'exhaustivité} = 1 - \frac{(Nb_O + Nb_E)}{N_t}$$

Dans notre cas, le taux d'exhaustivité est de $1 - ((2+2)/5) = 20\%$

Taux d'exhaustivité	Note sur 5
De 95 % à 100 %	5
De 90 % à 95 %	4
De 75 % à 90 %	3
De 50 % à 75 %	2
Taux < 50 %	1

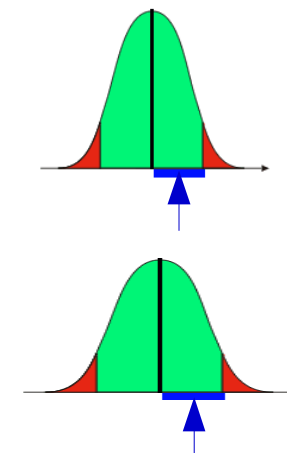
Dans notre cas, la note sera de 1



La précision thématique se définit comme la précision des attributs quantitatifs, la justesse des attributs non quantitatifs et la justesse de classement. Ce critère s'adresse, non plus aux objets eux-mêmes, mais aux informations qu'ils portent.

Attributs quantitatifs

L'incertitude de la valeur d'attribut avec un seuil de signification de 95 % est la demi-longueur de l'intervalle défini par une limite supérieure et inférieure, dans laquelle la valeur véritable se situe avec une probabilité de 95%

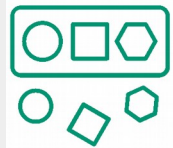


Attributs qualitatifs

Réalité terrain	leu de donn		
	habitation	bâti industriel ou commercial	bâ
Habitation	2152	25	
bati industriel ou commercial	185	651	
bati remarquable	2	13	
total	2339	689	

Cette **matrice de classement erroné (MCM)** peut être réduite au coefficient **kappa** qui est destiné à mesurer le degré d'accord d'affectation aux classes en supprimant les classements erronés.

Exemple de matrice de classement erroné des entités



La cohérence logique se définit comme le degré de cohérence interne des données selon des règles de modélisation et les règles inhérentes à la spécification de produit.

Cohérence conceptuelle

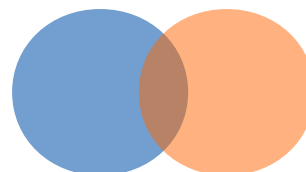
id	libelle	id_flore	<i>id_faune</i>
1	libelle1	1	
1	...	2	
2	...		
2	...	5	
3	libelle5	4	

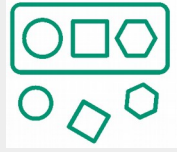
id_flore	flore
1	flore1
2	...
3	...
4	flore4

id_faune	faune
1	faune1
2	...
3	...
4	faune4

Non-unicité d'identifiants, clés étrangères absentes...

Chevauchements entre objets de classes différentes



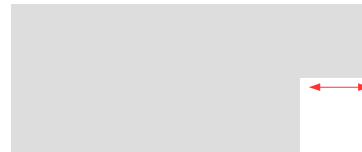


La cohérence logique se définit comme le degré de cohérence interne des données selon des règles de modélisation et les règles inhérentes à la spécification de produit.

Domaine de valeurs

Donnée

Domaine réel ou standard



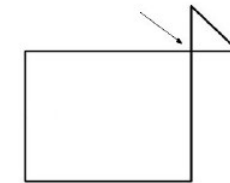
Cohérence de format

PRESCRIPTION_LIN.shp	29/09/2016 13:08	Fichier SHP
PRESCRIPTION_LIN.shx	29/09/2016 13:08	Fichier SHX
PRESCRIPTION_OBJETS_POLYGONES.dxf	31/05/2017 15:59	Fichier AutoCAD
PRESCRIPTION_PCT.dbf	29/09/2016 13:08	Classeur OpenOf
PRESCRIPTION_PCT.pri	29/09/2016 13:08	Fichier PRJ

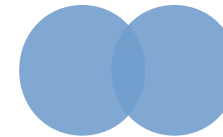
id_flore	flore
Type Float	Type Text
1	flore1
2	...
4	flore4

Cohérence topologique

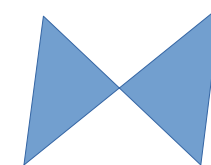
Micro-surfaces non valides



Auto-chevauchements, lacunes



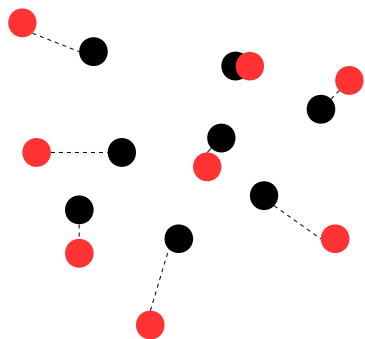
Auto-intersections





La précision de position se définit comme la précision de la position des entités au sein d'un système de référence spatial. Elle concerne aussi bien la position d'une entité prise isolément que sa position par rapport aux objets qui l'entourent.

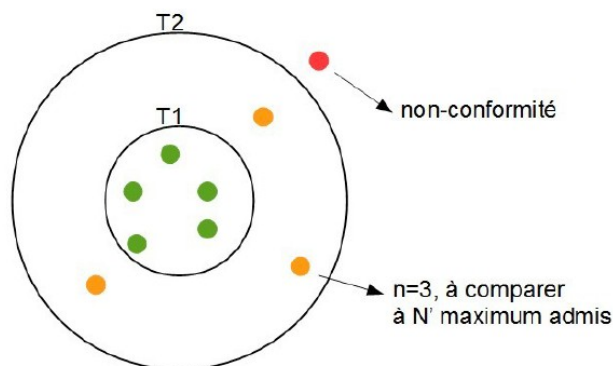
Objets ponctuels



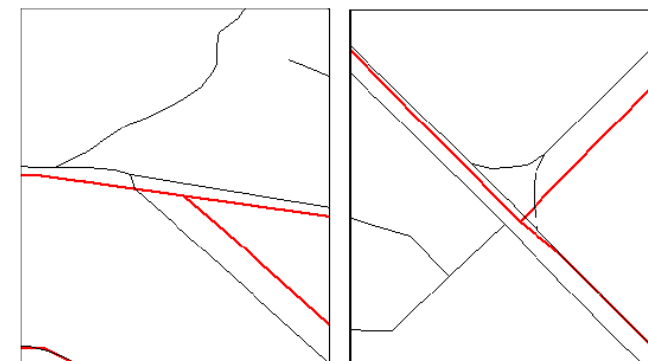
La valeur moyenne des incertitudes ou écart moyen en position

en 2 dimensions $e_i = \sqrt{(x_{mi} - x_{ti})^2 + (y_{mi} - y_{ti})^2}$

$$\bar{e} = \frac{1}{N} \left(\sum_{i=1}^N e_i \right)$$



Objets linéaires et surfaciques



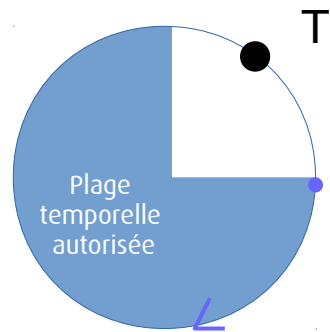
(en rouge, réseau à contrôler, en noir source de contrôle)

- Réduire la précision de position d'un réseau peut revenir à le réduire à la précision de position de **ses sommets**.
- On peut aussi relever, dans le jeu de données et dans la source de contrôle des points représentant des **changements de cap**.

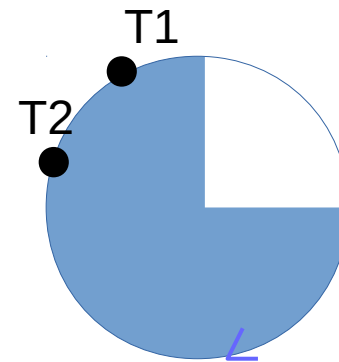


Ce critère s'intéresse à la dimension temporelle des attributs d'objets d'une base de données géographiques. Il en vérifie la précision temporelle ainsi que la cohérence logique et chronologique.

Cohérence vis-à-vis du domaine de valeurs



Cohérence chronologique



T1 fait référence à un évènement situé chronologiquement après T2, mais se trouve transcrit sous la forme d'une date postérieure dans la donnée

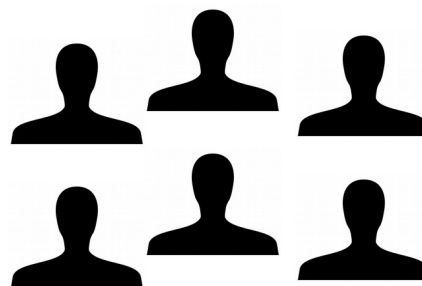
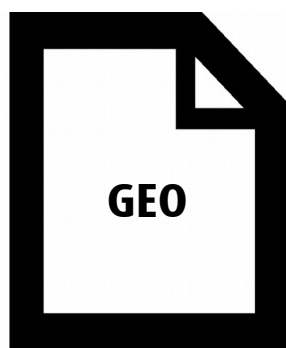
Exemple d'un PLU

LIBELLE	LIBELONG	TYPEZONE	VOCATION	NOMFIC	URLFIC	INSEE	DATAPPRO	DATVALID
Uh		U	01	[XXXXX]_reglemen t_20160620.pdf		[XXXXX]	20160620	20171103
Ur1		U	01	[XXXXX]_reglemen t_20160620.pdf		[XXXXX]	20181103	20140428

La qualité interne

Elle mesure la qualité d'un lot de données produit **selon des spécifications rédigées au préalable** (cahier des charges). Ces spécifications détaillent la structure et le contenu de la base de données.

Les méthodes de contrôle de la qualité interne sont décrites dans la norme ISO 19157.



La qualité externe

Elle se définit comme **l'adéquation à un besoin exprimé**, l'aptitude d'un jeu de données à satisfaire un usage donné, en particulier un type d'usage générique comme une exploitation cartographique, un usage à but statistique, ou la constitution d'un observatoire.

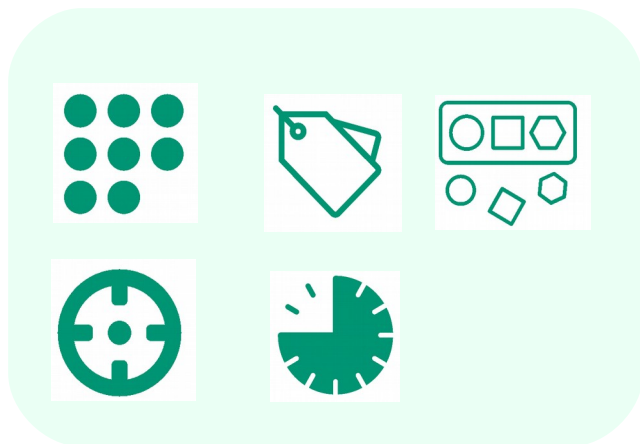
La qualité perçue pour un usage spécifique

Elle exprime la perception d'un utilisateur pour un besoin spécifique. Elle a été démocratisée avec **l'essor de l'Internet et la multiplication des systèmes de notation** par les communautés d'internautes. Il n'y a pas de raison que l'information géographique échappe à cette évolution.

Une méthode d'évaluation de la qualité des données décrit des procédures et/ou des traitements à appliquer aux données afin de parvenir à **un résultat de la mesure de la qualité**. Elle apporte **un éclairage complémentaire** aux résultats bruts du contrôle qualité.

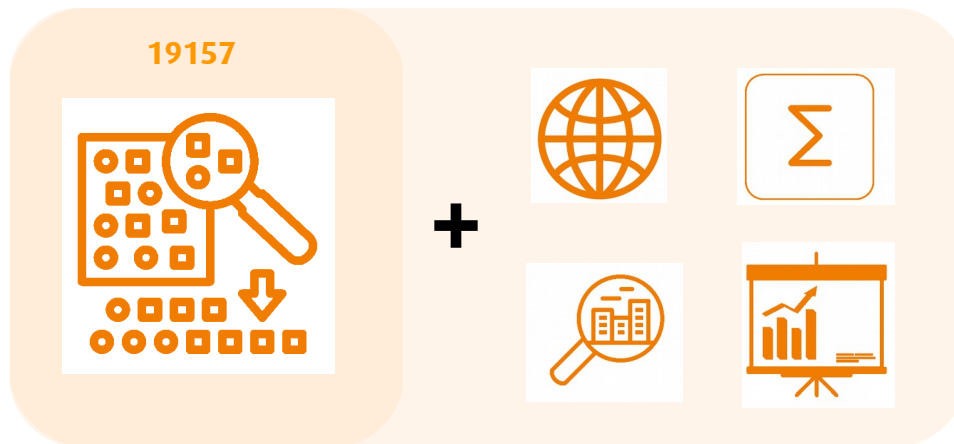
La norme ISO 19157 ne détaille pas les méthodes possibles d'évaluation, ce qui est pourtant une information essentielle pour quiconque se lance dans une démarche de qualification des données. Par contre, elle est très riche sur l'aspect **échantillonnage**.

Critères



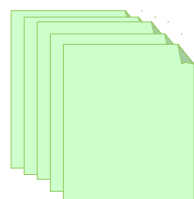
+

Méthodes



+

QuadoGéo et ses 10 fiches



Généralités



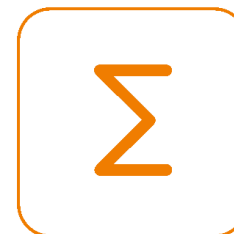
Concepts généraux sur la qualité des données géographiques

Contexte de contrôle



Connaître le cadre dans lequel est conduit l'évaluation de la qualité d'un jeu de données.
Se préparer à différentes situations.

Éléments statistiques



Connaître quelques éléments statistiques afin de fonder les mesures finales sur des procédés incontestables

Échantillonnage



Constituer des échantillons représentatifs en vue du contrôle qualité des données géographiques, et appliquer différentes méthodes d'échantillonnage spatial.

Méthodes de représentation

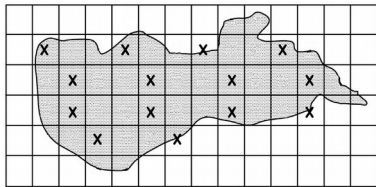
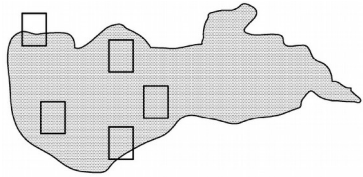


Proposer des méthodes de représentation, riches ou synthétiques, afin de partager une vision de la qualité mesurée.



Échantillonnage

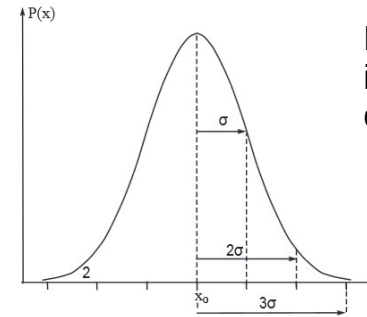
Exemple de deux méthodes d'échantillonnage : orienté surface et semi-aléatoire



Une méthode d'échantillonnage sera sélectionnée en fonction de l'homogénéité de la répartition spatiale et de l'homogénéité des classes d'objets



Éléments statistiques

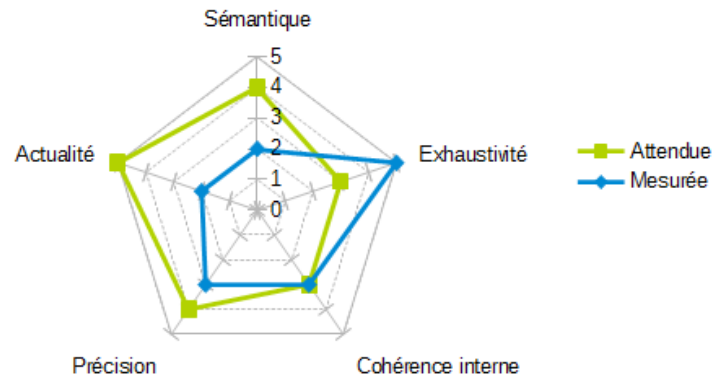


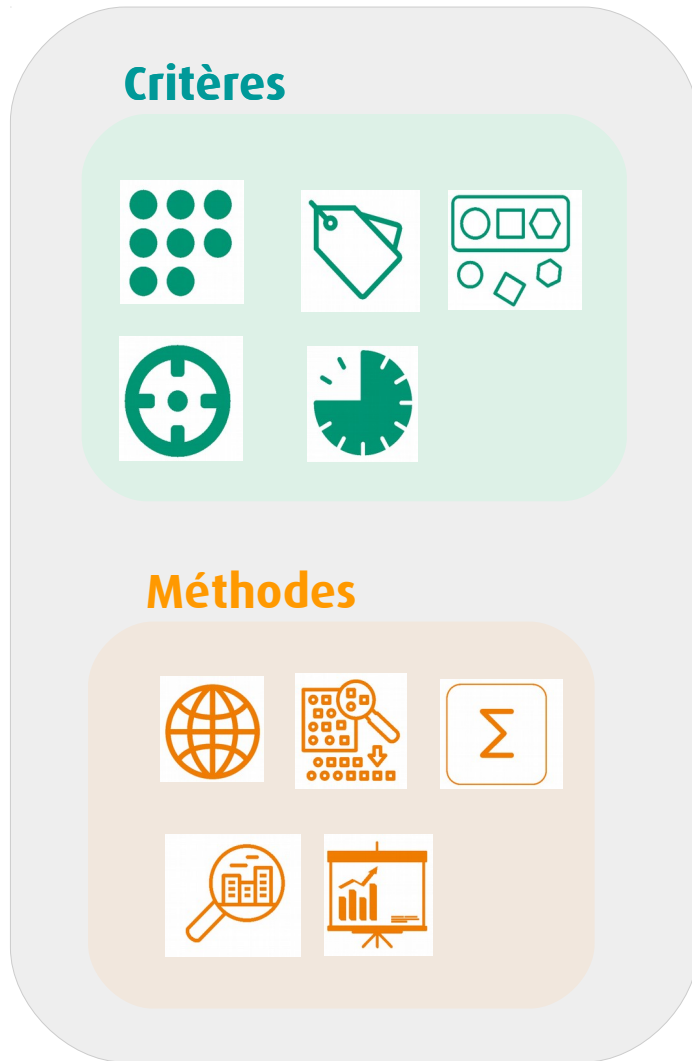
Moyennes, écart-types, intervalles et niveaux de confiance, incertitude



Méthodes de représentation

Représentation simplifiée des notes attendues et mesurées d'un jeu de données





Partager

Fournir une analyse de la norme et rendre **un contenu documentaire simplifié** par l'intermédiaire de [fiches](#) (projet QuaDoGéo), créer un nouveau portail d'(in)formation.

Former

Sur la base de cette documentation, construire [une mallette pédagogique](#) à partir de laquelle former des agents.

Outiller

Partager des **procédures**, des **méthodes** techniques basées sur certains logiciels pour l'évaluation de la Qualité, voire même concevoir **des logiciels « clé en main »** de rapportage et de visualisation des mesures.

Je vous remercie de votre attention

Cerema Méditerranée

Mardi 6 Février 2018- Séminaire « Outils et organisations pour la qualification des données géographiques »

Crédits (images Creative Commons issues de nounproject.org)

Carte *par Gira Park*

Enseignant *par Gregor Cresnar*

Entonnoir *par Kirill Ulitin*

Europe *par Anbileru Adaleru*

Fichier *par I cons*

Forêt *par Juan Pablo Bravo*

Globe *par Lloyd Humphreys*

Interface *par Aftergrind*

Outil *par Hea Poh Lin*

Utilisateur *par Edward Boatman*

Usine *par Abdul Karim*

Valise *par Shashank Singh*

Visuels des critères

Cohérence logique *par Made by Made*

Exhaustivité *par Made by Made*

Précision de position *par Gregor Cesnar*

Précision thématique *par Pictohaven*

Qualité temporelle *par Icon 54*

Visuels des méthodes

Contexte de contrôle *par Yu Luck*

Échantillonnage *par Becris*

Globe *par Proletkult Graphik*

Méthode de représentation *par NOPIXEL*

Statistiques *par Burak Kucukparmaksiz*