

 <p>Liberté • Égalité • Fraternité</p> <p>RÉPUBLIQUE FRANÇAISE</p>  <p>Conseil national de l'information géographique</p> <p>Commission règles & qualité</p> <p>Groupe de Travail Qualité des Données Géographiques (QuaDoGéo)</p>	<p>Groupe de Travail animé par :</p>  <p>Direction territoriale Ouest Département Villes et territoires</p>
<p>COMPTE-RENDU DE REUNION (Distanciel)</p>	
<p>Objet : Réunion #6 du GT QuaDoGéo du 5 juin 2020.</p>	

Ordre du jour :

1. Accueil et rappels
2. Validation du [précédent compte-rendu](#). Points d'info et d'actu
Retour sur deux mini-sessions du webinaire « La data dans les territoires »
(**S. Levêque**)
3. Retours d'expérience :
Qualification de la BD OCS GE : point de vue de l'utilisateur (**A-M. Clément**)
Qualification de la cohérence entre MOS et document d'urbanisme sur le
SCoT du Pays d'Arles (**S. Rolle**)
4. Synopsis pratiques et note méthodologique pour la qualification de
données
5. Restitution de la qualification des données (**M. Rajerison**)
6. Services de remontée de la qualité des données par les utilisateurs

Liste des participants

Animateurs : Arnaud Gallais (CEREMA)

Nom Prénom	Organisme	Présent	Excusé
Jean-Marie ARSAC	AZIMUT		x
Jean-Philippe AST	IGN / OCS GE NG	x	
Jean-Marie BOURGOGNE	OpenDataFrance / Validata		x
Jérôme BOUTET	Idéo Ternum Bourgogne-Franche-Comté		x
Romain BUCHAUT	CRIGE PACA		x
Anne-Marie CLEMENT	Métropole Nice Côte d'Azur	x	
Olivier DISSARD	CGDD/DRI/MIG		x
Arnauld GALLAIS	Cerema Ouest	x	
Stéphane GARCIA	IGN / Cellule normalisation	x	
Clément JAMET	Métropole de Lyon	x	
Mathieu Le Moal	Axes Conseil	x	
Hélène LEUFROY	Géo17	x	
Stéphane LEVEQUE	Cerema DTec TV	x	
Jocelyne MARC	IGN / Mission Qualité	x	
Suzanne NICEY	Idéo Ternum Bourgogne-Franche-Comté	x	
Nicolas PY	IGN Centre Est		x
Clémence RABEVOLO	IFREMER	x	
Mathieu RAJERISON	Cerema Méditerranée	x	
Stéphane ROLLE	CRIGE PACA	x	
Pierre VERGEZ	IGN / Mission CNIG	x	

Réunion #6 du GT QuaDoGéo du 5 juin 2020 (Distanciel).		
Mise en relecture du compte-rendu	Inconnu	Arnaud Gallais
Relecture du compte-rendu	Inconnu	Les participants.
Validation du compte-rendu	02/10/2020 (réunion #7)	Arnaud Gallais
Prochain rendez-vous le 2 octobre 2020 en distanciel.		

1. Accueil et rappels

2. Validation du précédent CR - Points d'info et d'actu

Le compte-rendu du précédent GT CNIG QuaDoGéo ([lien](#), [permalien](#)) est relu et validé.

La session test de la FOAD pour des collectivités est reportée au 2^{ème} semestre 2020.

A. Gallais a étudié avec le BRGM les pistes d'améliorations du Registre français pour les "Métadonnées relatives à la qualité des données géographiques", ce qui a bien permis d'introduire des hyperliens au sein des entrées du registre, mais l'affichage par défaut suivant l'ordre des items de la norme au sein de chaque entrée du registre n'a pas pu être obtenu (il faut actuellement trier suivant l'item « notation »). Le registre sera présenté à la prochaine réunion de la Commission Règles de mise en œuvre, et est actuellement publié sur la page de ressources du GT CNIG QuaDoGéo sous forme de projet.

Les livrables BAL et Equipco ont été harmonisés et la note méthodologique générale pour la qualification de données a été élaborée (*cf. infra*).

L'action de développement « d'un outil simple (macro excel ou autre...) de saisie des évaluations et de génération du résultat graphique conforme à cette maquette » est reportée à la prochaine réunion.

S. Levêque fait un retour sur les webinaires "data dans les territoires" développés en partenariat avec OpenDataFrance et le CRIGE PACA. Ces 8 webinaires d'une heure ont vocation à sensibiliser sur la culture de la donnée, l'opendata, les besoins, pratiques et possibilités de collectivités en la matière. Il s'agit d'une formation qui s'adresse aux agents de collectivités désireux d'acquérir les bons réflexes pour prendre les bonnes décisions dans la mise en œuvre d'un projet d'ouverture des données. Les deux premières sessions ont proposé trois modules gratuits pendant la période de confinement. Elles ont rencontré un franc succès (80 inscrits en quelques jours...) et ont obtenu une excellente évaluation. Plus d'information ici : <https://www.cerema.fr/fr/actualites/webinaires-data-territoires> ([permalien](#)).

3. Qualification de la BD OCS GE : point de vue l'utilisateur

A-M. Clément présente la qualification de la BD OCS GE sur le territoire de la Métropole Nice Côte-d'Azur (MNCA). Présentation disponible [ici](#) ([permalien](#)).

En préalable, **A-M. Clément** rappelle l'obligation réglementaire faite aux Collectivités Territoriales d'intégrer le développement durable dans les outils de planification et d'aménagement du territoire (*article [L101-2](#) du code de l'urbanisme*). Afin de répondre à ces obligations MNCA souhaite disposer d'une OCS GE précise, respectueuse de ses spécifications et complète.

L'OCSGE produite par l'IGN sur le territoire de MNCA répond globalement aux utilisations d'aménagement de l'espace métropolitain ; d'élaboration du PLU métropolitain (PLUm) ;

de protection et mise en valeur de l'environnement et la politique du cadre de vie ; de gestion des services d'intérêt collectif ; etc. Différents besoins sont pris en compte tels le suivi des indicateurs d'évaluation du PLUm concernant par exemple la trame verte et bleue, la consommation des espaces, la vulnérabilité des personnes et des biens face aux risques, et la mesure de l'imperméabilité des sols.

Comparée au MOS préexistant, la BD OCSGE donne de très bons résultats avec toutefois des réserves marquées sur l'unité minimale d'intérêt (UMI) de 2500 m² jugée inapplicable aux espaces d'intérêt environnemental télédéetectables du Haut-Pays niçois, tels : les zones humides, les ouvrages de protection, une part importante (19%) des lacs et la quasi-totalité de cours d'eau. Les zones de risques naturels sont insuffisamment identifiées. De même, l'imperméabilité des sols est sous-estimée car tous les bâtis ne sont pas pris en compte, en particulier les 98 ha cumulés de piscines sur le territoire de la Métropole. De plus, les ouvrages de protection : digues, jetées, brise-lames ne sont pas toujours recensés malgré leur réelle utilité pour le territoire.

Afin d'améliorer la qualification du bâti, MNCA a créé un cinquième niveau de qualification du bâti comprenant : les villages de montagne perchés ou au creux d'un vallon (CS1.1.1.1.1), Les zones pavillonnaires (CS1.1.1.1.2), les grands ensembles (CS1.1.1.1.3), la ville constituée (CS1.1.1.1.4).

En conclusion de ce retour d'expérience, **A-M. Clément** indique que la BD OCSGE doit être considérée comme un socle performant destiné à être :

- complété de nombreuses adaptations de façon à ce tout ce qui est télédéetectable puisse s'y retrouver, telles les zones humides et les cours d'eau en raison de leur valeur environnementale ;
- amélioré par un ajustement de la valeur de l'unité minimale d'intérêt (UMI) aux spécificités du territoire afin de mieux prendre en compte les besoins liés à la gestion des risques naturels et de l'artificialisation des sols.

Ainsi, pour ce qui concerne les ouvrages de protection, MNCA a demandé la création d'un poste de couverture dédié et le classement de ces ouvrages en poste de couverture anthropisée plutôt qu'en zone naturelle.

Pour ce qui concerne les piscines MNCA souhaite que tous les bâtis jointifs au bâti principal (terrasses, piscines) soient retenus comme bâtis ou non-bâtis imperméables (en lien avec l'identification exhaustive de l'artificialisation des sols).

A. Gallais remarque que la BD OCSGE correspond à ses spécifications de production et de qualité, et note un besoin pour MNCA de disposer de bases de données complémentaires (BD Topage pour les cours d'eau, etc.) pour répondre plus finement à certains des besoins exprimés.

H. Leufroy demande si cette évaluation qualitative du produit s'est accompagnée d'une évaluation quantitative via une recette du produit. **A-M. Clément** confirme que la recette a en effet porté sur plus de 7000 entités polygonales d'occupation du sol.

Faisant le lien avec l'[observatoire des données nationales sur l'artificialisation des sols](https://artificialisation.developpement-durable.gouv.fr/)¹, **S. Levêque** remarque l'influence de l'unité minimale d'intérêt (UMI) ou de collecte (UMC) sur la mesure de l'artificialisation.

¹ En décembre 2024, redirigé sur <https://artificialisation.developpement-durable.gouv.fr/> ([permalien](#)).

J-P. Ast précise les apports de l'IA dans le nouveau processus de production de la BD OCSGE, dont un prototype est en cours sur le territoire d'Arcachon avant une production prévue à partir de 2021 par tiers du territoire chaque année. Il souligne que si, par exemple, les piscines peuvent être mieux détectées par l'intelligence artificielle, elles n'en demeurent pas moins hors spécifications actuelles du produit. Si le processus de production « nouvelle génération » est automatisé, les spécifications de produit BD OCSGE ne sont pas modifiées.

4. Qualification de la cohérence entre MOS et document d'urbanisme sur le SCoT du Pays d'Arles

S. Rolle présente un travail d'évaluation rapide du PLU de la commune d'Arles à partir du MOS du Pays d'Arles. Travail mené par des étudiants sur deux lots de données de même millésime (2017). Présentation disponible [ici \(permalien\)](#).

Il rappelle que le MOS décrit un état objectif de l'espace, tandis que le PLU est un document de planification décrivant un projet politique d'aménagement et d'urbanisme. Les deux documents sont donc fortement corrélés mais avec des différences.

La qualification du critère de précision thématique a fait l'objet d'une matrice de classement erroné basée sur la conformité de classement des zones AUc, AUs, et AU du PLU par rapport au MOS 2017. La matrice de confusion indique que 12 % des zonages A (agricoles) du PLU correspondent à des zones naturelles (N) dans le MOS.

Le MOS avait lui-même préalablement fait l'objet d'une évaluation du critère de précision de position indiquant des erreurs géométriques pouvant potentiellement générer des erreurs de désignation et/ou de segmentation de l'espace.

S. Rolle indique à quel point cette courte étude est emblématique par rapport à la problématique de la qualification des données :

- elle constitue un mini-test avec de premiers résultats concrets quant aux réalisations du groupe de travail, [série de fiches Cerema - qualifier les données géographiques](#), etc.
- elle confirme les besoins de disposer :
- d'une méthodologie et d'un processus de qualification (*cf infra « note méthodologique »*)
- d'un outillage générique pour calculer les indicateurs (matrices, calculs automatisables, etc.)
- d'une plateforme de contrôle géométrique à définir : méthode Qgis / GEOS, OpenJump, postgis, basé sur l'excellent rapport « [Vérification et correction des géométries](#) » ([permalien](#)) publié sur [GéOInformations](#).
- Enfin, cette étude met en évidence le besoin d'une méthode de restitution et représentation de la qualité / qualification d'un lot de données.

Le groupe de travail s'interroge sur ses moyens pour initier la réalisation de tels outils...

M. Le Moal et **C. Jamet** indiquent que les outils de contrôle de données font appel à des algorithmes distincts et que des contrôles opérés par divers outils sur un même lot de données peuvent donner des résultats différents.

Les outils open source présentent l'avantage d'utiliser des algorithmes ouverts et documentés ce qui n'est pas toujours le cas des outils propriétaires. Ils citent à titre d'exemple des [outils de contrôle des données au format citygml](#) pour vérifier les primitives géométriques en 3D.

5. Synopsis pratiques pour la qualification de données

*Par **S. Levêque** et **S. Rolle**.*

Il s'agit de construire une méthodologie accompagnée d'un synopsis pratique permettant de déterminer, en fonction de la nature d'une donnée, quels critères et quelles mesures sont à utiliser et si possible dans quel ordre. Un objectif sous-jacent est d'évaluer la qualité des données en fonction de l'usage afin d'approcher une évaluation de la qualité externe des données.

***J-M. Arzac** a initié l'élaboration de la nomenclature des usages génériques de l'information géographique. Cette nomenclature, qui doit être extensible, permettra de placer les priorités et pondération des critères et mesures de qualité en face des usages. Afin d'alimenter la réflexion globale sur la méthode, des expérimentations pratiques ont été réalisées par deux groupes sur les thématiques « Base Adresse locale » et « Equipements collectifs publics ».*

L'action est finalisée pour les deux volets Adresse et Equipements collectifs publics.

A l'instar de ce que **S. Levêque** avait présenté à la [précédente réunion](#) pour ce qui concerne le cas de la base adresse locale, **S. Rolle** présente les trois livrables finalisés sur le cas des équipements collectifs : le [synopsis \(permalien\)](#), [l'arbre de décision \(permalien\)](#) pour l'évaluation des critères de qualification, ainsi que sa [notice d'accompagnement \(permalien\)](#).

Il suggère de mener une réflexion quant à l'utilisation de formats spécifiques pour certains champs. Pris en exemple, le standard [OpeningHour](#) permet de formater les horaires d'ouverture et de fermeture d'un lieu. L'utilisation de tels standards présente l'avantage de grandement faciliter l'exécution d'algorithmes de contrôles maintenus par une communauté. [OpeningHour](#) participerait en l'occurrence à valider la cohérence temporelle.

S. Levêque rappelle le besoin d'harmoniser les méthodes d'évaluation de la qualité pour différentes thématiques et de multiples données. Il a synthétisé et détaillé cette démarche dans la [fiche méthodologique \(permalien\)](#) en cours d'élaboration, qu'il présente en séance. Elle se veut d'un abord concret et facile, avec des illustrations par l'exemple.

La méthodologie actuelle consiste à s'interroger sur ce que l'on qualifie, puis pour quelle raison, ce qui mène directement à la question de(s) usage(s) envisagé(s) à partir des données évaluées.

Cette question des usages amène à s'interroger sur l'existence d'un « registre officiel des usages de l'information géographiques » ou, dans la négative, le moyen d'étoffer la nomenclature initialisée par **J-M. Arzac**. La liste des thèmes INSPIRE est évoquée, mais elle reste majoritairement liée au domaine de l'environnement et on n'y retrouve pas tous les thèmes actuellement listés au §2.1 de la note méthodologique.

Cette note évoque également les étapes de hiérarchisation des critères qualité, et de représentation des résultats de l'évaluation, comme synthèse des travaux des réunions précédentes.

S. Garcia indique qu'il manquerait dans le déroulement méthodologique actuel l'étape de réalisation des évaluations et/ou mesures qualité.

C. Jamet questionne l'éventuel choix d'une représentation des résultats de la qualification selon l' (les) usage(s).

Le groupe de travail est d'avis qu'il sera pertinent d'inclure cette fiche méthodologique dans le cursus de la formation à distance sur la qualification des données géographiques.

Décision/actions :

- Abonder la nomenclature des usages génériques de l'information géographique
- Relire et commenter la note méthodologique (tous)

6. Restitution de la qualification des données géographiques

Par S. Levêque. [La présentation de la session précédente est ici.](#)

Les diagrammes polaires offrent une excellente solution de restitution graphique de la qualité d'un jeu de données géographiques, à condition :

- de les distinguer suivant l'usage technique attendu de la donnée ;
- et qu'ils portent sur leurs branches les mesures correspondant aux différents critères de la qualité de l'information géographique.

A l'instar de sites web du style « Les numériques » le groupe de travail s'oriente vers :

- deux notes globales : la note du testeur (vert), et celle des utilisateurs (orange) ;
- une rubrique « déposer un avis » ;
- des sous-notes suivant des thèmes (ex : facilité d'obtention, facilité d'usage, « fraîcheur », etc.) ;
- une rubrique « points forts / points faibles », potentiellement transposable en : « usages recommandés / usages non recommandés ».

Le développement d'un outil simple (macro excel ou autre...) de saisie des évaluations et de génération du résultat graphique conforme à cette maquette n'ayant pas progressé pendant la période du confinement, le point sur la restitution de la qualification des données est reporté à la prochaine réunion.

Décision/actions :

- Développer un outil simple (macro excel ou autre...) de saisie des évaluations et de génération du résultat graphique conforme à cette maquette. Idéation **S. Levêque** a priori sur la base du design graphique proposé aux précédentes réunions, développement **M. Rajerison** avec une personne de Lyon Métropole.

7. Services de remontée de la qualité des données par les utilisateurs

Par **M. Rajerison**. La présentation est [ici \(permalien\)](#), elle fait suite aux « [Aventures du Da\(ta\)hu](#) » (nov. 2019) ([permalien](#)).

Les retours et avis des utilisateurs sont essentiels pour améliorer la qualité des données ouvertes mais **M. Rajerison** constate le manque de filière officielle structurée pour la remontée de leurs avis sur l'utilisation et l'évaluation de la qualité des lots de données. Ceci a pour effet de tourner les utilisateurs vers les réseaux sociaux, fréquemment Twitter.

Le standard [Geospatial User Feedback](#) (GUF), déjà présenté (cf. [CR de septembre 2019](#)), reste complexe et finalement peu utilisé par les plateformes locales et/ou portails grand public.

Après avoir présenté plusieurs exemples de remontée d'informations (datagouv, datasud, DataGrandLyon) **M. Rajerison** relaie une démarche canadienne basée sur l'[opendata catalogue](#) de la ville de Toronto : le Data quality score (DQS) : « Towards a Data Quality Score in open data » (voir sa présentation [partie 1 \(permalien\)](#) et [partie 2 \(permalien\)](#)), et

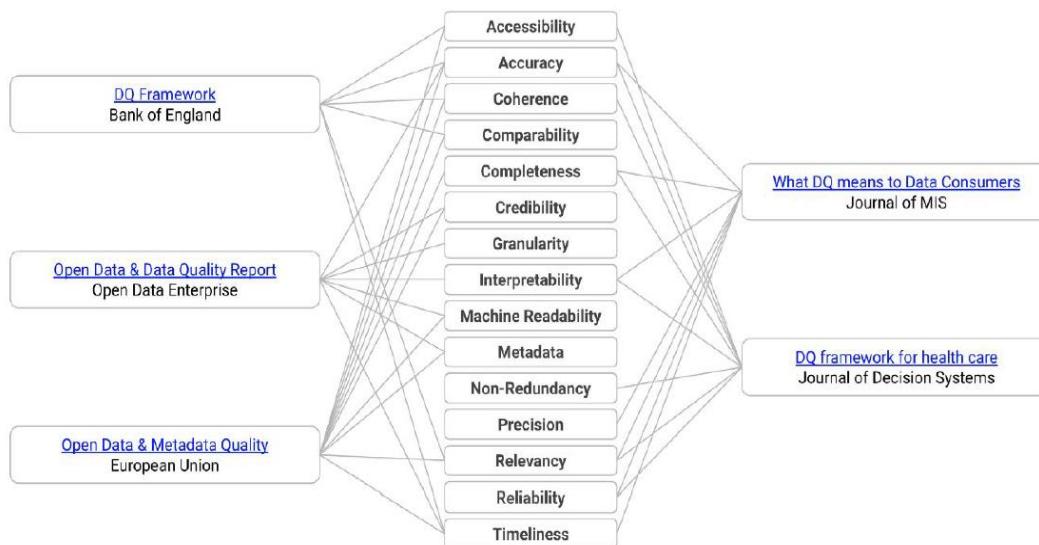
l'exemple des données sur l'emplacement des [stationnements vélo](#)). Le DQS applique le principe qu'« *il vaut mieux peu de données de qualité que beaucoup de données de faible qualité* ». Dans cet esprit, les utilisateurs disposent de moyens pour qualifier les lots de données ,selon une échelle de 1 à 7, suivant des critères de : description (via les métadonnées), granularité, compréhension, usage, fraîcheur, accessibilité et exhaustivité de l'information.

	1	2	3	4	5	6	7
METADATA: Is the data well described?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
GRANULARITY: How atomic is the data?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
INTERPRETABILITY: How easy is it to understand the data?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
USABILITY: How easy is it to work with the data?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
FRESHNESS: How close to creation time is the data published?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
ACCESSIBILITY: Is the data easy to access?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
COMPLETENESS: How much data is missing?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Pour autant, ce portail de données reconnaît volontiers que « *malgré l'importance de leur qualité, les données ouvertes ne bénéficient pas réellement d'une véritable évaluation car ses acteurs ne sont pas des experts en la matière* » dans la mesure où ils ne disposent en particulier pas des moyens d'évaluer les critères d'exactitude, cohérence, précision, exhaustivité, etc. critères désignés par la norme ISO19157.

Ces deux approches complémentaires traduisent une nette différence entre l'évaluation de la qualité externe (capacité de répondre à un ou des usages, cf le « sondage ci-dessus ») et la qualité interne d'un lot de données, car elles ne présentent à l'évidence pas les mêmes besoins et exigences quant aux critères qualité.

De plus, l'évaluation de la qualité externe peut en fonction de la thématique ou de l'usage d'un lot de données (que l'on se place du point de vue d'une banque, d'un système d'information, ou d'un utilisateur avec un besoin précis, etc.), faire appel à des ensembles de critères très différents :



M. Rajerison fait au sujet de la « remontée de l’avis de l’utilisateur » le parallèle avec l’univers du développement logiciel, déjà bien rodé en la matière. Il prend en exemple le process de remontée d’erreurs mis en place sur des plateformes de développement telles que [Github](#), dont il conviendrait de s’inspirer en identifiant en particulier les analogies entre la qualification des données et celle des logiciels.

Par exemple, la remonté des anomalies sous forme de « issues » sur le produit OpenAdresses offre la possibilité de « taguer » ces anomalies suivant différents critères.

C. Jamet indique que le portail DataGandLyon offre également un tel champ libre textuel qui de plus référence le jeu de données.

M. Rajerison conclut en constatant que les processus de remontée d’informations sur la qualité s’avèrent encore assez « libres » au sens où ils ne sont guère contraints par la structuration ou la normalisation. DQS et Github lui paraissent être d’excellentes sources d’inspiration quant aux processus de remontée d’avis des utilisateurs. Il recommande de poursuivre en procédant à une expérimentation sur une plateforme de données ouvertes.

S. Garcia remarque que la plateforme data.gouv.fr moissonne un très important volume de lots de données et suggère qu’un service de remontée des avis des utilisateurs lui soit connecté.

A. Gallais suggère une future présentation par l’IGN de son processus de remontée d’anomalies via son [espace collaboratif](#).

Avancement :

A ce stade, la production effective ou en cours du GT CNIG QuaDoGéo comprend :

- Les [fiches méthodologiques](#) du Cerema ;
- Le [registre français de mesures de qualité des données géographiques](#) ;
- La [carte mentale](#) au sujet de la qualification des données suivant la norme ISO 19157 ;
- Deux pistes exploratoires de qualification de données (synopsis, logigramme) sur des thèmes particuliers ;
- La [méthodologie](#) de qualification de données (*à consolider...*), comprenant :
 - une nomenclature des usages génériques de l'information géographique ;
 - une maquette de restitution de la qualité des données ;
- Une [présentation du GUF](#), outil de retour utilisateur quant à la qualité des données.

Prochaine réunion (distanciel) le 2 octobre 2020.