Descriptif du fichier Referentiel_Pays-et-territoires.csv en date du 22 juin 2023

Présentation

Cette liste des pays et territoires du monde a été établie au premier semestre de l'année 2023. Elle a pour ambition de mettre en relation les principaux référentiels de même nature publiés en langue française par des organismes publics français de référence (code officiel géographique de l'INSEE) et des institutions internationales (ONU, ISO, Union internationale des télécommunications). Sont considérés comme États souverains les entités reconnues comme telles par la France. Les autres entités sont mentionnées comme relevant d'un État souverain si c'est le cas. Certaines entités ne sont rattachées à aucun État souverain lorsque leur statut ou leur tutelle est contesté ou indéfini.

La liste ainsi établie ne définit pas un partitionnement politique ou géographique du monde. Outre les contestations territoriales qui entretiennent l'incertitude quant au rattachement administratif de certaines emprises, certains territoires peuvent se référer à plusieurs entrées différentes de la liste. En effet, les référentiels pris en compte peuvent qualifier de plusieurs noms distincts un même territoire. Ainsi, les Provinces espagnoles d'Afrique (nom apparaissant dans le code officiel géographique de l'INSEE) peuvent être considérées comme regroupant les Canaries et les Présides.

Lorsque, dans les dénominations de référence, le libellé d'un territoire comprend plusieurs noms géographiques (toponymes) séparés par « et » ou par une virgule, chacun de ces noms correspond à une entrée distincte de la liste, à l'exception des entités pour lesquelles la conjonction de coordination est intégrée au toponyme (par exemple, Saint-Pierre-et-Miquelon), ce qui se traduit par la présence de traits d'union de part et d'autre de la conjonction de coordination.

Les discordances entre référentiels entraînent la duplication de certains intitulés ainsi que la présence de valeurs vides.

Les caractères sont encodés au format Windows-1252

Le séparateur des champs est le point-virgule.

Liste et description des champs

ID : chaîne de 6 caractères. Ce champ permet de disposer d'un identifiant unique par appellation géographique de territoire ou de pays. Hormis le Vatican, qui dispose de deux enregistrements (correspondant chacun à l'un des deux indicatifs téléphoniques attribués par l'UIT à ce pays) avec le même ID, chaque enregistrement dispose d'un ID unique. Par construction, celui se compose d'un préfixe sur deux caractères proposant une répartition géographique des pays et territoires par continent (AF pour Afrique, AM pour Amérique, AS pour Asie, EU pour Europe, OC pour Océanie et Antarctique), d'un caractère de liaison (_) et d'un suffixe composé de 3 chiffres.

NOM_ALPHA : Chaîne de caractères. Nom « court » de l'entité. Lorsqu'il s'agit d'un État reconnu par la France, la graphie est celle recommandée par le texte N° 63 du JORF du 21 avril 2019 (NOR : CTNR1910801K), actualisant le texte N° 91 du 24 septembre 2008 (NOR : CTNX0818389X). Pour les

autres entités et territoires dépendants, la référence est le document « Entités géopolitiques souveraines ou dépendantes » publié par la commission nationale de toponymie (CNT/CNIG) le 4 décembre 2015. Les corrections et compléments apportés proviennent du code officiel géographique de l'INSEE.

Pour les toponymes comportant un terme générique (Îles, ...), un article (La) ou une description non signifiante sur le plan géographique (République démocratique du ...), cet élément est rejeté entre parenthèses après le premier terme géographique. Cela permet un classement alphabétique par « nom géographique signifiant ».

NOM_COURT : Chaîne de caractères. Nom « court » de l'entité. Lorsqu'il s'agit d'un État reconnu par la France, la graphie est celle recommandée par le texte N° 63 du JORF du 21 avril 2019 (NOR : CTNR1910801K), actualisant le texte N° 91 du 24 septembre 2008 (NOR : CTNX0818389X). Pour les autres entités et territoires dépendants, la référence est le document « Entités géopolitiques souveraines ou dépendantes » publié par la commission nationale de toponymie (CNT/CNIG) le 4 décembre 2015. Les corrections et compléments apportés proviennent du code officiel géographique de l'INSEE.

Sont conservées entre parenthèses les désignations levant les éventuelles ambiguïtés (partie française ...).

ARTICLE : Chaîne de caractères, éventuellement vide. Ce champ permet d'indiquer l'article utilisé pour l'appellation complète du pays, dans sa forme courte, dans le cadre d'un texte. Lorsqu'il s'agit d'un État reconnu par la France, la graphie est celle recommandée par le texte N° 63 du JORF du 21 avril 2019 (NOR : CTNR1910801K), actualisant le texte N° 91 du 24 septembre 2008 (NOR : CTNX0818389X). Pour les autres entités et territoires dépendants, la référence est le document « Entités géopolitiques souveraines ou dépendantes » publié par la commission nationale de toponymie (CNT/CNIG) le 4 décembre 2015.

ISO_ETAT : Chaîne de 2 caractères, éventuellement vide. Ce champ permet d'indiquer l'État reconnu par la France disposant de la souveraineté sur l'emprise décrite par le nom géographique. Il est vide pour les territoires dont le statut ou la souveraineté est contesté ou non défini. Lorsqu'il est rempli, il coïncide avec le code alphanumérique sur 2 caractères défini par la norme ISO 3166-1 alpha-2 telle que publiée par l'ISO début 2023.

NOM_LONG_ETAT : Chaîne de caractères, éventuellement vide. Ce champ permet d'indiquer la forme longue du nom de l'État reconnu par la France disposant de la souveraineté sur l'emprise décrite par le nom géographique. Il est vide pour les territoires dont le statut ou la souveraineté est contesté ou non défini. Lorsqu'il est rempli, la graphie est celle recommandée par le texte N° 63 du JORF du 21 avril 2019 (NOR : CTNR1910801K), actualisant le texte N° 91 du 24 septembre 2008 (NOR : CTNX0818389X).

ISO_3166_NOM : Chaîne de caractères, éventuellement vide. Ce champ indique la forme courte en langue française du nom de l'entité selon la norme ISO 3166-1 publiée par l'ISO début 2023. Il est vide pour les territoires non référencés par cette norme. Ce libellé peut correspondre à une juxtaposition de noms géographiques réunis par « et » ou par une virgule.

IND_TEL_UIT : Chaîne 0 à 3 chiffres. Ce champ mentionne pour chaque territoire l'indicatif téléphonique qui lui est attribué selon l'annexe au bulletin d'exploitation n°1114 du 15 décembre 2016 de l'Union internationale des télécommunications. Il est vide pour les territoires non

mentionnés par ce référentiel. En revanche, deux indicatifs différents étant associés au Vatican (379 et 39), celui-ci apparaît à deux reprises dans la liste publiée.

ISO_CODE_NUM : Chaîne de 3 chiffres, éventuellement vide. Ce champ indique le code numérique attribué au territoire par la norme ISO 3166-1 publiée par l'ISO début 2023. Il est vide pour les territoires non référencés par cette norme.

ISO_alpha3 : Chaîne de 3 caractères, éventuellement vide. Ce champ indique le code numérique attribué au territoire par la norme ISO 3166-1 publiée par l'ISO début 2023. Il est vide pour les territoires non référencés par cette norme.

ISO_alpha2 : Chaîne de 3 chiffres, éventuellement vide. Ce champ indique le code numérique attribué au territoire par la norme ISO 3166-1 publiée par l'ISO début 2023. Il est vide pour les territoires non référencés par cette norme.

SOUVERAIN : 1 caractère (O ou N) qui précise si le territoire est reconnu comme souverain par la France (O) ou non (N).

NOM_ONU : Chaîne de caractères, éventuellement vide. Ce champ mentionne les graphies en français des territoires publiées par l'ONU via son département des affaires économiques et sociales, division de statistique (<https://unstats.un.org/unsd/methodology/m49/>) début 2023. Ce champ est vide pour les territoires non référencés par ce répertoire. Le libellé peut correspondre à une juxtaposition de noms géographiques réunis par « et » ou par une virgule.

CODE_COG : Chaîne de 5 chiffres, éventuellement vide. Ce champ reprend le code officiel géographique du pays ou du territoire selon le référentiel de l'INSEE début 2023 (<https://www.insee.fr/fr/information/2028273>). Les codes périmés correspondant à des entités n'ayant plus de signification dans la géographie politique actuelle n'ont pas été retenus (Tchécoslovaquie, RDA, Turquie d'Europe...). De même, les territoires répertoriés par l'INSEE comme ne disposant pas d'un code officiel géographique spécifique ne sont pris en compte que s'ils apparaissent dans d'autres référentiels (Macao, Jersey...) et s'ils ne relèvent pas de la France.

Le champ n'est pas rempli pour la France et les emprises sous souveraineté française car un code officiel géographique est attribué à chaque commune française ou entité administrative assimilée à une commune. Il n'est pas rempli pour les territoires non mentionnés dans ce référentiel (Antarctique, Sercq, Îles mineures éloignées des États-Unis).