

Normes des formats de fichiers ICARTT V1.1

Statut du mémo

Ce document RFC (« Request for comments ») décrit les normes pour les formats de fichiers ICARTT, les implémentations et les ressources disponibles.

La distribution de ce mémo est illimitée.

Changements apportés aux explications :

Mise à jour de V1 à V1.1 :

Section 2.1.B. Informations sur la localisation – Ajout de clarifications à propos de la précision des mesures. « Il est recommandé de reporter la latitude et la longitude décimales à la précision maximale de l'instrument. Pour les instruments GPS classiques en aviation, la latitude et la longitude devraient être reportées au moins jusqu'à la cinquième décimale. Il est recommandé de rapporter l'altitude en mètres. »

Section 2.6. Logiciel de balayage de fichiers - Correction d'une erreur grammaticale dans la première phrase.

Avis de droit d'auteur

Copyright © Gouvernement des États-Unis tel que représenté par l'administrateur de la National Aeronautics and Space Administration. (2013). Tous droits réservés.

Résumé

Les normes des formats de fichiers ICARTT ont été développées pour remplir les besoins en matière de gestion de données pendant la campagne 2004 du « International Consortium for Atmospheric Research on Transport and Transformation » (ICARTT). L'étude ICARTT a consisté en 11 expériences de terrain individuelles hautement coordonnées impliquant plus de 300 agences gouvernementales et universités participantes en provenance de 5 pays, c'est-à-dire les États-Unis, le Canada, le Royaume-Uni, l'Allemagne et la France. Un format commun et simple d'utilisation des fichiers de données, le format de fichier ICARTT, a été établi pour cette étude dans le but premier de faciliter l'échange de données et de promouvoir les collaborations entre les différentes équipes scientifiques afin d'atteindre les objectifs scientifiques du projet ICARTT. Le fichier ICARTT est un fichier format texte qui se compose d'une section d'en-tête (métadonnées) contenant les informations descriptives essentielles (p.ex. l'origine des données, les incertitudes, les informations de contact et une brève présentation des méthodes de mesure) et d'une section contenant les données. Bien qu'il ait été initialement conçu pour les données collectées par voie aérienne, le format de fichier ICARTT s'est révélé pratique pour d'autres études mobiles et au sol, et pour des types de données variés. Au-delà du succès de l'étude ICARTT, le format de fichier ICARTT est maintenant largement accepté par la communauté scientifique dans le domaine de la recherche atmosphérique sur le terrain et a été récemment utilisé dans des études aériennes majeures financées par la NASA, la NSF, la NOAA et des partenaires internationaux.

Tables des matières

STATUT DU MÉMO	1
AVIS DE DROIT D’AUTEUR	1
RÉSUMÉ	2
TABLES DES MATIÈRES	3
1. INTRODUCTION – ORIGINE DU FORMAT STANDARD DE FICHIER ICARTT	4
2. SPECIFICATION DU FORMAT DU FICHIER	5
2.1.A. INFORMATIONS SUR LE TEMPS	6
2.1.B. INFORMATIONS SUR L’EMPLACEMENT GÉOGRAPHIQUE.....	6
2.1.C. MESURES	7
2.2. NOM DES FICHIERS	8
2.3. SPECIFICATIONS DU FORMAT DE FICHIER POUR LES SERIES DE DONNEES TEMPORELLES ICARTT	10
2.3.A. STRUCTURE	10
2.3.B. INFORMATIONS DE LA SECTION D’ENTÊTE D’UN FICHIER	11
2.3.C EXEMPLES.....	14
2.4 SPECIFICITES DE FORMAT POUR LES FICHIERS ICARTT MULTIDIMENSIONNELS ..	16
2.4.A STRUCTURE	16
2.4.B. EXEMPLES.....	17
2.5 FORMATS DE FICHIERS POUR DES DONNEES AERIENNES NON STANDARD	19
2.6. LOGICIELS DE BALAYAGE DE FICHIER	20
3. BIBLIOGRAPHIE	21
4. ADRESSE DES AUTEURS	21
5. ANNEXE A	22

1. Introduction – Origine du format standard de fichier ICARTT

Depuis le début des années 80, la NASA et ses agences partenaires ont dirigé plus de 30 campagnes de terrain majeures dans la troposphère, pour étudier la composition atmosphérique d'un large éventail de régions géographiques. En comparaison des données satellites, ces données aériennes fournissent une perspective historique plus longue, un plus grand jeu d'espèces/ de paramètres observés et une résolution spatiale, verticale comme horizontale, plus importante. En conséquence, les observations aéroportées sont d'une valeur unique pour la communauté des modélisateurs afin d'évaluer l'habilité de cette discipline à prédire la composition atmosphérique future et ses impacts sur les changements climatiques et la qualité de l'air. De plus, les observations aériennes peuvent aussi être utilisées pour développer/ améliorer les données *a priori* utilisées dans les algorithmes d'extraction satellitaires. Néanmoins, l'utilisation des données aéroportées dans l'évaluation et la validation des modèles présente des défis significatifs. Parmi ces défis se trouve le manque d'uniformisation des données. Les mesures aériennes existantes sont archivées sous des formats variés, ce qui rend leur utilisation et leur échange difficile. L'établissement de protocoles pour la construction de fichiers de données standardisés est une étape importante pour faciliter l'échange de données entre les différentes communautés scientifiques.

L'étude de terrain de l'International Consortium for Atmospheric Research on Transport and Transformation (ICARTT) a été menée pendant l'été 2014. Celle-ci consistait en 11 expériences de terrain, indépendantes mais hautement coordonnées, p.ex. les études NASA INTEX-NA, NOAA NEAQS-ITCT et UE ITOP. Alors que chacune de ses études de terrain (financées par des agences gouvernementales issues de 5 pays de chaque côté de l'Atlantique Nord, c.-à-d. les États-Unis, le Canada, le Royaume-Uni, l'Allemagne et la France) avait des objectifs scientifiques concentrés sur sa région d'étude, celles-ci partageaient un objectif scientifique principal commun : examiner les processus clés reliés aux émissions d'aérosols et des précurseurs de l'ozone, ainsi que leurs transformations chimiques et leur élimination pendant leur transport vers et au-dessus de l'océan Atlantique Nord. La campagne ICARTT a impliqué plusieurs centaines de participants, de multiples plateformes, aérienne et navales, mais aussi des infrastructures au sol. Un large volume de données a été généré par les équipes de recherche individuelles au sein de chaque expérience de terrain correspondante. Un format de fichier de données commun, le format ICARTT, a été créé pour faciliter le partage de données entre les différentes équipes scientifiques et pour remplir les besoins en matière de gestion de données durant toutes les phases de l'étude de terrain, p.ex. le déploiement sur le terrain, le traitement et l'analyse des données post-déploiement et les publications.

Le fichier ICARTT est un fichier texte auto-descriptif relativement simple d'utilisation. Le format du fichier a été construit à partir de deux formats de données aéroportées bien établis : NASA Ames et GTE. Comme ses prédécesseurs, le format de fichier ICARTT est conçu pour le traitement de données issues de mesures aériennes *in situ* mais possède des capacités limitées pour intégrer des données issues de la télédétection aérienne ou au sol (p.ex. lidar), d'autres mesures prises au sol et certains aspects des données satellitaires. Le format ICARTT se compose de deux sections : une section d'en-tête (métadonnées) et une section de données. La section d'en-tête

possède les instructions pour extraire des données à partir du fichier et les informations essentielles relatives à la description des données (p.ex. l'origine des données, les informations de contact, une brève présentation des méthodes de mesures, les incertitudes liées aux mesures et les commentaires de révision des données) afin que l'utilisateur ait suffisamment d'informations, soit pour utiliser directement les données soit pour contacter le chercheur principal en charge des mesures afin d'obtenir des précisions à propos de certains problèmes.

À cause de la campagne de terrain ICARTT, le format des fichiers de données ICARTT a été exposé à un large éventail de chercheurs travaillant en milieu aérien. Le succès de la campagne ICARTT a naturellement conduit à une acceptation encore plus grande du format de fichier ICARTT dans les dernières études aériennes. Par exemple, le format de fichier ICARTT a été adopté pour les campagnes de terrain NASA INTEX-B et NSF MILAGRO en 2006. Plus récemment, l'étude de terrain POLARCAT de l'année polaire internationale a utilisé le format de fichier ICARTT comme standard pour les programmes participants financés par la NASA, la NOAA, et ses partenaires internationaux en France et en Allemagne. Son acception grandissante et son utilisation répandue, particulièrement dans le domaine des mesures aériennes *in situ*, ont propulsé le format de fichier de données ICARTT au rang de référence pour la communauté des études aériennes.

Ce document définit les normes du format de fichier ICARTT dans la section 2, laquelle inclut les spécifications du format, les conventions d'appellation, les spécifications des sections d'en-tête et les applications aux divers types de données. Plusieurs exemples sont donnés afin de fournir des précisions supplémentaires. Une brève description du logiciel de balayage qui peut être utilisé pour tester la compatibilité des fichiers de données avec les normes du format de fichier ICARTT, est également donnée ici.

2. Spécification du format du fichier

À l'occasion d'une vaste campagne d'étude aérienne comme celle d'ICARTT 2004, de nombreux types de données sont collectés. Beaucoup d'ensembles de données sont de simples séries temporelles avec un ou plusieurs paramètres mesurés séquentiellement (et simultanément) dans le temps. Cependant, certains ensembles de données sont multidimensionnels, en cela que les observations prises par un instrument à un moment donné dans le temps sont distribuées spatialement. Les données du profileur de vent en sont un exemple : les échantillons, moyennés sur 30 minutes pour une période de temps donnée, sont regroupés en informations d'altitude, tandis que la vitesse et la direction du vent, ainsi que la température, sont fournies pour chaque altitude. Un autre exemple, plus extrême, est la sortie des modèles tridimensionnels. Des données comme celles-ci ne peuvent certainement pas être représentées comme une seule série temporelle. Les sections 2.1 – 2.2 ci-dessous présentent le format ICARTT pour différents types de données, l'emphase étant mise sur les données de type séries temporelles standard, lesquelles sont typiques des mesures aériennes *in situ*. Les sections 2.3 – 2.4 sont spécifiques aux données de type séries temporelles standard. La section 2.5 offre des indications sur les séries de données temporelles non standard.

Le format de fichier ICARTT n'a pas de restriction sur le nombre de caractères par ligne ou le nombre de caractères par enregistrement. Le nom du fichier est limité à 127 caractères en longueur. Comme ces prédécesseurs, le format ICARTT utilisera aussi l'ensemble de caractères ASCII pour la construction de fichiers. La section des données du fichier ne sera composée que de caractères numériques de type ASCII incluant les notations scientifiques, les virgules comme séparateurs de champs et les espaces dans le but d'améliorer la clarté visuelle (alignement) des données.

Le caractère fin de ligne pour les fichiers textes diffère en fonction du système d'exploitation informatique. De nombreux outils modernes de traitement de texte traitent et convertissent le caractère fin de ligne automatiquement si bien que ce problème passe inaperçu pour la majorité des utilisateurs. Néanmoins, c'est un problème que certains utilisateurs vont rencontrer et les gestionnaires de données devraient en avoir conscience. Il existe de nombreuses ressources sur internet (p.ex. Wikipédia) traitant de ces problèmes en détails et proposant des solutions pour les résoudre. Une solution rapide peut être aussi simple que d'utiliser le « mode ASCII » pour les transferts ftp de fichiers ou utiliser un éditeur de texte différent (plus récent, p.ex. WordPad), etc. Le logiciel de balayage mentionné (plus bas) traite automatiquement les caractères de fin de ligne lorsque nécessaire.

2.1.A. Informations sur le temps

Les fichiers de données sont tenus de rapporter le temps de début et de fin de chaque mesure. C'est la seule façon non ambiguë de représenter le temps d'intégration des mesures, exception faite des données collectées de façon continue à 1 Hz ou moins pouvant être représentées par une seule étiquette de temps. Le temps doit être reporté en secondes UTC à partir du début de la journée (date) à laquelle la mesure a débuté. Cette date apparaît dans l'en-tête du fichier comme dans son nom. Le temps rapporté devrait augmenter de façon monotone, même lors du passage au jour suivant. Pour les mesures en continu, la chronologie rapportée doit être ininterrompue entre la première et la dernière mesure effectuée. Cela peut être accompli en utilisant des identificateurs de données manquantes pour tenir compte des lacunes dans les données, causées par la calibration ou d'autres périodes d'arrêt des instruments. Deux exceptions importantes pour la continuité chronologique existent : les données avec espacement ou temps d'intégration irrégulier(s) et les données satellitaires qui peuvent contenir d'importantes lacunes dans les données à cause d'interférences avec les nuages, de changements de mode, etc. Malgré ces lacunes, il faut tenir compte du large volume de données collecté par ces plateformes (voir section 2.5 ci-après).

2.1.B. Informations sur l'emplacement géographique

Tous les points de données doivent être associés à une latitude, une longitude et une altitude. La latitude et la longitude devraient être rapportées en degrés décimaux avec les latitudes sud et les longitudes ouest représentées par des nombres négatifs (c.-à-d. pas d'identifiant N, E, O, S). Il est recommandé de rapporter la latitude et la longitude décimales à la précision maximale de l'instrument. Pour les instruments GPS classiques en aviation, la latitude et la longitude devraient être rapportées au moins jusqu'à la cinquième décimale. Il est recommandé de rapporter l'altitude en mètres. L'altitude doit être définie clairement compte tenu des nombreux types de mesures d'altitude utilisés (altitude pression, altitude GPS, altitude géo potentielle, altitude radar, etc.). Il

est souvent avantageux de rapporter les informations sur la position géographique (et autre) dans un fichier indépendant (p.ex. un fichier regroupant les paramètres relatifs à un avion). Dans ce cas, il n'est pas nécessaire de répéter l'information pour chaque instrument de la plateforme dans les fichiers de données. Les fichiers de données peuvent plutôt se référer au fichier de paramètres dans la section de l'entête. Cette option est spécifiée plus bas à la section 2.3.B.

2.1.C. Mesures

En général, chaque fichier contient les données relatives à un paramètre ou une espèce. De multiples variables par fichier sont autorisées seulement si celles-ci ont toutes été mesurées exactement sur la même base temporelle, comme, par exemple, à l'aide du même instrument. La représentation numérique d'une variable est définie par les unités dans lesquelles elle a été mesurée. Le format ICARTT contient le nécessaire pour appliquer un facteur d'échelle aux données. Cependant, il est recommandé de fixer tous les facteurs d'échelle à 1, à moins que cela ne soit très contraignant. Si de très grands, ou très petits nombres sont requis, ils peuvent alors être représentés en notation exponentielle, comme 1.01E9 ou 5.23E-6.

i. Incertitudes

L'incertitude de mesure est associée de façon inhérente à chaque mesure. Le format de données ICARTT nécessite de rapporter l'incertitude TOTALE afin d'inclure tous les effets, systématiques et aléatoires. Si les estimés d'incertitude sont disponibles pour chaque période de mesure, les incertitudes peuvent faire l'objet d'une tabulation dans une colonne séparée immédiatement après celle des données dans le fichier. Cependant, cette condition peut être assouplie si les incertitudes peuvent être données en tant qu'information dans la section de l'entête du fichier. Par exemple, si toutes les incertitudes peuvent être calculées par une fonction ayant comme entrée n'importe quel point des données, alors cette formule peut être incluse comme information d'entête. Il est impératif que l'intervalle de confiance sigma (p.ex. 1 sigma ou 2 sigma) soit rapporté avec les incertitudes. De même, les unités des incertitudes doivent être explicitement présentées dans l'entête du fichier. Lorsque l'incertitude absolue est donnée, la même unité doit être utilisée pour l'incertitude et pour sa mesure associée. Pour l'incertitude relative, la valeur devrait être rapportée en pourcentage (p.ex. 30% ou 10%).

ii. Données manquantes

Les données manquantes sont simplement – manquantes, c'est-à-dire que l'instrument n'a pas enregistré de données à cause d'une calibration ou d'un problème de fonctionnement. Les valeurs manquantes sont représentées par des nombres négatifs assez grands pour ne pas être considérés comme des données réelles. Pour le format de fichier ICARTT, la valeur est -9999 (ou -99999, etc.). D'autre part, des données en-dessous ou au-dessus de la limite de détection (« limit of detection » ou LOD) ne sont pas réellement « manquantes » puisqu'elles fournissent certaines informations. Alors que certains chercheurs choisissent de tabuler toutes leurs

données quantifiables, incluant les valeurs négatives pour les concentrations, d'autres choisissent de représenter ces points comme des valeurs inférieures à une limite de mesure quantifiable. Un traitement similaire est également effectué pour des données possédant des valeurs plus grandes que la LOD supérieure. Ces conditions sont indiquées par deux étiquettes que l'on substitue aux valeurs des données manquantes. L'étiquette pour des valeurs SUPÉRIEURES à certaines LOD SUPÉRIEURES (« Upper Limit Of Detection » ou ULOD) est -7777 (ou -77777, etc.) et l'étiquette pour des valeurs INFÉRIEURES à certaines LOD INFÉRIEURES (« Lower Limit Of Detection » ou LLOD) est -8888 (ou -88888, etc.). Ces étiquettes (si elles sont utilisées) et les valeurs des LOD supérieures et inférieures sont documentées à des emplacements spécifiques dans la section d'entête du fichier (voir plus bas). Si les valeurs des LLOD et ULOD varient d'un point à l'autre, elles doivent être fournies dans une colonne séparée des données.

iii. Caractères de délimitation des données

Les virgules sont utilisées pour délimiter les champs de données au sein des enregistrements (lignes) de données dans un fichier.

2.2. Nom des fichiers

Les caractéristiques propres aux différentes conventions de dénomination des fichiers ont été adaptées ici. Les noms de fichiers pour le format ICARTT, limités à 127 caractères ou moins, sont définis comme suit :

donnéesID_localisationID_AAAAMMJJ[hh[mm[ss]]]_R#[_L#][_V#][_commentaires].ict,

Où les seuls caractères autorisés sont a-z A-Z 0-9 _ . - (c'est-à-dire, caractères alphanumériques majuscules et minuscules, trait de soulignement, point et trait d'union). Tous les champs qui ne sont pas entre crochets sont requis. Les champs sont décrits comme suit :

donnéesID : courte description des paramètres/ espèces mesurées, instrument ou modèle (p.ex. O3 ; RH ; VOC ; PTRMS ; MM5)

localisationID : courte description du site, de la station, plateforme, laboratoire ou institut

AAAA : année à quatre chiffres

MM : mois à deux chiffres

JJ : jour à deux chiffres

hh : heure à deux chiffres, optionnel

mm : minute à deux chiffres, optionnel

ss : seconde à deux chiffre, optionnel

R : numéro de révision des données

L : numéro de lancement, optionnel

V : numéro du volume, optionnel

commentaires : informations additionnelles, optionnel

extension : **ict**, extension du fichier, toujours « ict »

Le trait de soulignement est **UNIQUEMENT** utilisé pour séparer les champs du nom du fichier ; il possède une signification précise pour le logiciel de balayage (voir section 2.6). Pour séparer les caractères à l'intérieur d'un champ pour une question de lisibilité, utilisez des lettres minuscules et majuscules. L'utilisation du trait d'union, bien qu'autorisée, est déconseillée car l'utilisation de ce caractère dans le nom du fichier peut causer des problèmes avec certains systèmes d'exploitation ou des logiciels réseau plus anciens. Les crochets « [] » contiennent les paramètres optionnels mais ils n'apparaissent pas dans le nom du fichier. Les dates et les heures dans les noms de fichiers sont toujours au format UTC. La date et l'heure dans le nom du fichier donnent la date/l'heure à laquelle les données rapportées commencent dans le fichier (fichiers de données), ou la date/l'heure à laquelle l'image a été prise (fichiers d'images). Pour les fichiers de données relatifs aux avions et aux sondes, la date réfère toujours à la date UT du lancement.

Le champ **donnéesID** est une courte chaîne de caractères utilisée pour identifier les paramètres du fichier. Pour les fichiers contenant une ou deux variables, le nom des variables peut être utilisé dans le nom du fichier. Pour les fichiers dans lesquels sont représentés de nombreuses variables, il est préférable d'indiquer une classe de composés dans le nom du fichier (p.ex., COV ; taux de photolyse) ou l'abréviation de l'instrument utilisé pour faire les mesures (p.ex. PTRMS).

Le champ **localisationID** est utilisé pour identifier la plateforme de mesure, le site, la station ou la source (laboratoire ou institut) de l'information comprise dans le fichier de données. Quelques exemples seraient DC8, BAE146, RHBrown, GOME (satellite), IoS (site d'Appledore Island), ChebPt (site de Chebogue Point) et autres. Il peut être utile d'avoir un ensemble standardisé d'abréviations utilisé pour une mission de terrain donnée. L'équipe scientifique de la mission devrait décider elle-même de ces abréviations.

Le paramètre **R** n'est pas optionnel dans le format de données ICARTT. Chacun doit spécifier un code de révision des données qui retrace les différentes mises à jour effectuées. Cela nécessite également que la documentation liée à ces mises à jour (p.ex. nouvelle calibration, erreur dans le déroulement des opérations, etc.) soit rapportée dans l'entête du fichier (voir section 2.3.B). Pour cela, nous spécifions un numéro de révision « _R# » où le trait de soulignement est un élément exigé pour séparer les champs (nécessaire pour certains logiciels de balayage de fichiers). Le numéro de révision « # » doit correspondre à celui spécifié dans la section des commentaires de l'entête du fichier (voir section 2.3.B).

Les paramètres optionnels «_L#» et «_V#» peuvent être nécessaires dans certains cas particuliers. Si le contenu du fichier se rapporte à un second ou un troisième lancement d'aéronef à la date indiquée, alors le numéro de décompte du lancement «_L#» (c.-à-d. L2, L3, etc.) doit apparaître après l'identifiant «R» mais avant le numéro de volume, si existant (voir plus bas). Le numéro de lancement L1 est assumé lorsque «_L#» n'apparaît pas dans le nom du fichier. Si le fichier de données correspond à un volume appartenant à un ensemble de données multi-volumes, alors un numéro de décompte du volume «_V#» (c.-à-d. V1, V2, V3, etc.) doit apparaître après le paramètre «R» (et le paramètre «L», si présent) et être séparé par un trait de soulignement du reste de l'identifiant. Le numéro du volume (le «#» dans «V#») doit correspondre au numéro de volume donné dans la section de l'entête. Lorsque le paramètre «_V#» n'apparaît pas dans le nom du fichier, on assume que l'ensemble de données est contenu dans un seul volume.

Le paramètre des commentaires optionnels permet d'ajouter des informations requises par le chercheur principal (ou le gestionnaire des données) pour identifier le contenu du fichier, mais cela ne s'intègre pas dans les autres champs du nom de fichier. Cette option doit être utilisée avec parcimonie.

2.3. Spécifications du format de fichier pour les séries de données temporelles ICARTT

2.3.A. Structure

Le format de fichier ICARTT pour les séries de données temporelles est structuré de façon à imiter le format de fichier Ames, **File Format Index (FFI) = 1001**. La définition du FFI dans le format Ames est la suivante : *le File Format Index (FFI) est utilisé pour définir de façon unique le format du fichier d'échange. Par référence aux options de format prédéfinies, la valeur du FFI détermine le nombre de variables INDÉPENDANTES, que les valeurs des variables INDÉPENDANTES et dépendantes soient numériques ou de simples chaînes de caractères, le format de l'entête du fichier et le format des enregistrements de données.*

Lorsque cela est possible, nous recommandons d'adopter un format de fichiers, pour les séries de données temporelles, conforme au format de fichier **FFI = 1001**.

FFI = 1001 : une variable indépendante, non limitée et réelle ; les variables primaires sont réelles ; il n'y a pas de variable auxiliaire ; les variables indépendantes et les variables primaires sont enregistrées dans le même document.

Cela indique qu'il y a une variable indépendante, habituellement le temps de départ, et que toutes les autres données dépendent de la variable indépendante. Dans les cas typiques, la variable fondamentale est le temps de départ de la mesure et les autres variables peuvent être définies telles que présentées dans l'exemple ci-dessous, où le nom des variables se réfèrent aux colonnes du fichier de données :

temps de départ

temps de fin

temps de mi-parcours
latitude
longitude
altitude/ élévation
données variable 1
incertitude variable 1
données variable 2
incertitude variable 2
< etc. >

Ce format est valable pour la plupart des séries temporelles de données mesurées n'importe quand, sous n'importe quelle période d'intégration arbitraire, et à n'importe quel endroit à la surface ou à proximité de la planète. Ce format peut aussi être condensé. Par exemple, si les mesures sont rapportées en continu à un intervalle d'une seconde ou moins, alors le temps de fin et le temps de mi-parcours n'ont pas besoin d'être inclus. De même, si les mesures sont prises à un emplacement fixe, alors la latitude, la longitude et l'élévation sont fixes et ces données sont incluses dans les informations de l'entête (voir section 2.3.B). Tel que souligné plus tôt, si la localisation des données (latitude, ect.) est incluse dans un fichier indépendant, alors ces colonnes peuvent être exclues, sous condition que le nom du fichier de localisation soit fourni dans les informations de l'entête du fichier de données. De la même façon, si l'incertitude est définie comme une fonction qui reste la même pour tous les points, alors celle-ci peut être incluse dans les informations d'entête et l'utilisateur peut alors calculer lui-même les incertitudes.

2.3.B. Informations de la section d'entête d'un fichier

Pour le format des données ICARTT, des informations additionnelles sont requises et doivent être incluses dans les sections des commentaires. L'entête la plus commune est montrée plus bas comme un exemple ; des entêtes plus spécifiques sont décrites comme des modifications de la forme commune. Les séparateurs de champs (/d'éléments) sont les virgules uniquement. Pour des séparateurs délimitant le texte au sein d'un élément, utilisez les traits de soulignement. L'ordre dans lequel apparaissent les données dans l'entête est présenté plus bas. Les mots apparaissant en gras doivent apparaître dans l'entête, suivis de l'information pertinente. Des exemples pertinents d'entête sont fournis à la suite de cette liste.

- Nombre de lignes de l'entête, index du format de fichier (FFI, la plupart des fichiers utilisent 1001) – délimité par une virgule.

- Nom de famille du chercheur principal, prénom/ initiales.
- Institution/ affiliation du chercheur principal.
- Description de l'origine des données (p.ex. le nom de l'instrument, de la plateforme ou du modèle, etc.).
- Nom de la mission (habituellement l'acronyme).
- Numéro de volume du fichier, nombre total de volumes (ces valeurs entières sont utilisées lorsque les données nécessitent plus d'un fichier par jour ; pour les données ne nécessitant qu'un seul fichier, ces valeurs sont fixées à 1,1) – délimité par une virgule.
- Date UTC du moment auquel commencent les données, date UTC de la réduction ou de la révision – délimité par une virgule (aaaa, mm, jj, aaaa, mm, jj).
- Intervalle des données (cette valeur décrit l'intervalle de temps (en secondes) entre deux enregistrements de données consécutifs. C'est l'intervalle (constant) entre les valeurs de la variable indépendante. Pour les données à 1 Hz, l'intervalle de données est 1 ; pour les données à 10 Hz, sa valeur est de 0.1. Tous les intervalles plus longs qu'une seconde doivent être rapportés comme des temps de début et de fin d'enregistrement des données et la valeur de l'intervalle des données est fixée à 0. Le temps de mi-parcours est requis lorsqu'il ne représente pas la moyenne entre le temps de début et le temps de fin. Pour des informations additionnelles, voir la section 2.5 plus bas).
- Description ou nom de la variable indépendante (C'est le nom choisi pour le temps de départ. Il se réfère toujours au nombre de secondes UTC écoulées depuis le début de la journée à laquelle les mesures ont commencé. Il est à noter que la variable indépendante doit augmenter de façon monotone même lors du passage à un second jour.).
- Nombre de variables (valeur entière donnant le nombre de variables dépendantes : le nombre total de colonnes de données est cette valeur +1.).
- Facteur d'échelle (1 dans la plupart des cas, excepté lorsque ce n'est pas opportun) – délimité par une virgule.
- Indicateurs de données manquantes (soit -9999 (ou -99999, etc.) pour n'importe quelle condition de donnée manquante, à l'exception de la variable principale du temps (indépendante), qui n'est jamais manquante) – délimité par une virgule.
- Nom des variables et unités (Des noms courts des variables et des unités sont requis, puis le nom descriptif, plus long et optionnel peut suivre dans cet ordre ; délimités par des virgules. Si la variable n'a pas d'unité, entrez le mot clé « none » pour ses unités. Chaque nom court de variable et les unités (et les noms optionnels longs) sont entrés dans une seule ligne. Le nom court de la variable doit correspondre exactement au nom utilisé pour cette

variable dans l'entête de la colonne, c.-à-d. à la dernière ligne d'entête avant le début des données.).

- Nombre de lignes de commentaires SPÉCIAUX (valeur entière indiquant le nombre de lignes de commentaires spéciaux, n'incluant PAS cette ligne).
- Commentaires spéciaux (notes relatives aux problèmes ou circonstances se rapportant uniquement à ce document. Un exemple serait des commentaires/ problèmes associés à un vol en particulier.).
- Nombre de commentaires (c.-à-d. le nombre de lignes additionnelles fournissant des informations D'APPUI: valeur entière indiquant le nombre de lignes d'informations additionnelles, n'incluant PAS cette ligne.).
- Commentaires (informations D'APPUI : À cet endroit, les investigateurs peuvent décrire plus en détails les données et les paramètres de mesure. La structure des informations d'appui est décrite plus bas comme une liste de paires mots clés : valeurs. Spécifiquement, incluez ici les informations à propos de la plateforme utilisée, la géolocalisation des données, les méthodes de mesure et les commentaires de révision des données. Il est à noter que les informations concernant l'incertitude, la limite de détection supérieure (ULOD) et la limite de détection inférieure (LLOD) pour chaque variable ne sont pas optionnelles. Les valeurs ULOD et LLOD sont les valeurs, dans la même unité que les mesures, qui correspondent respectivement aux étiquettes -7777 (ou -77777, etc.) et -8888 (ou -88888, etc.) dans les données. La dernière ligne de cette section doit contenir tous les noms « courts » des variables, sur une seule ligne. Les mots clés dans cette section sont écrits en **GRAS** ci-dessous et doivent apparaître dans cette section de l'entête, avec les données correspondantes listées après la colonne. Pour les mots clés pour lesquels l'information n'est pas nécessaire ou non applicable, entrer la mention N/A.).

Le logiciel de balayage cherche les mots clés suivants (insensible à la casse) :

PI_CONTACT_INFO : numéro de téléphone, adresse postale et adresse courriel et/ou numéro de fax du chercheur principal.

PLATFORM : informations à propos de la plateforme ou du site.

LOCATION : incluant la lat/long/élev si applicable.

ASSOCIATED_DATA : nom des fichiers avec les données associées : données de localisation, paramètres de l'avion, données de navigation, etc.

INSTRUMENT_INFO : description des instruments, méthodes d'échantillonnage et particularités, références bibliographiques, etc.

DATA_INFO : unités et autres informations relatives à la manipulation des données.

UNCERTAINTY : les informations relatives aux incertitudes ; si l'incertitude est une valeur constante ou une fonction, dans les cas où elle n'est pas présentée comme une variable séparée.

ULOD_FLAG : -7777 (étiquette de la limite de détection supérieure, toujours -7 (ou -77, -777, etc.)).

ULOD_VALUE : valeur (ou fonction) de la limite de détection supérieure correspondant à l'étiquette -7777 dans l'enregistrement des données.

LLOD_FLAG : -8888 (étiquette de la limite de détection inférieure, toujours -8 (ou -88, -888, etc.)).

LLOD_VALUE : valeur (ou fonction) de la limite de détection inférieure correspondant à l'étiquette -8888 dans l'enregistrement des données.

DM_CONTACT_INFO : Gestionnaire des données – nom, affiliation, numéro de téléphone, adresse postale, adresse courriel et/ou numéro de fax.

PROJECT_INFO : date de début et de fin de l'étude, liens vers les sites internet, etc.

STIPULATION_ON_USE : (auto-explicatif).

OTHER_COMMENTS : toute autre information pertinente.

REVISION : #R, voir la discussion sur le nom des fichiers.

R# : commentaires spécifiques à une révision des données. Le numéro de révision et les commentaires associés sont **cumulatifs** dans le fichier de données, de façon à pouvoir retracer les changements appliqués aux données au cours du temps. Ajoutez les informations dans cette section de façon à ce que le dernier numéro de révision et les commentaires associés soient toujours présentés au début de cette partie de l'entête. Les données révisées les plus récentes doivent correspondre à la date de révision apparaissant à la ligne 7 de l'entête principale.

Indep_Var, VarName_1, VarName_2, VarName_3,...VarName_n

Pour les fichiers FFI = 1001, la formule permettant de calculer le nombre total de lignes de l'entête est : 14 + (# de variables dépendantes, données à la ligne 10) + (# de lignes de commentaires spéciaux) + (# de lignes de commentaires).

2.3.C Exemples

Deux exemples de séries de données temporelles (similaires), utilisant différentes formes d'informations d'entête, sont présentés ci-dessous. Il est à noter que le retour à la ligne

automatique dans les logiciels de traitements de texte peut donner l'impression d'un plus grand nombre de lignes de texte qu'il n'en existe en réalité. Dans les exemples qui suivent, la suite des lignes du dessus a été mise en retrait pour plus de clarté.

Exemple 1. Toutes les colonnes de données requises sont explicitement présentées.

Nom du fichier : HOX_DC8_20040712_R0.ict

```
36, 1001
Brune, William
Penn State University
ATHOS - OH and HO2 concentrations using cryo water mix ratio data for quenching corrections
ICARTT_INTEX
1, 1
2004, 07, 12, 2005, 01, 12
0
Start.UTC, seconds
4
1, 1, 1, 1
-9999, -9999, -9999, -9999
Stop.UTC, seconds
Mid.UTC, seconds
OH_pptv, pptv
HO2_pptv, pptv
0
18
PI_CONTACT_INFO: Address: 503 Walker Building, University Park, PA 16802; email: brune@essc.psu.edu;
PLATFORM: NASA DFRC DC8 - sampling underneath aircraft forward cargo bay location
LOCATION: Aircraft location data in nav_dc8_20040712_R0.ict file
ASSOCIATED_DATA: see ftp://ftp-air.larc.nasa.gov/pub-air/INTEXNA/
INSTRUMENT_INFO: OH/HO2 LIF
DATA_INFO: Units are pptv.
UNCERTAINTY: The absolute accuracy is conservatively estimated to be +/- 32% at two sigma confidence
ULOD_FLAG: -7777
ULOD_VALUE: N/A
LLOD_FLAG: -8888
LLOD_VALUE: N/A
DM_CONTACT_INFO: Bob Leshner; Penn State University; blesher@psu.edu
PROJECT_INFO: INTEX Mission 26 June-14 August 2004; California, Illinois, and New Hampshire
STIPULATIONS_ON_USE: Use of these data requires prior approval from William Brune
OTHER_COMMENTS: N/A
REVISION: R0
R0: Final Data
Start.UTC, Stop.UTC, Mid.UTC, OH_pptv, HO2_pptv
55526, 55545, 55535, 0.171, 9.791
55546, 55565, 55555, 0.180, 9.218
55566, 55585, 55575, 0.186, 9.767
55586, 55605, 55595, 0.176, 9.996
55606, 55625, 55615, 0.192, 9.513
55626, 55645, 55635, 0.185, 9.798
55646, 55665, 55655, 0.160, 9.834
```

Exemple 2. Toutes les colonnes de données requises sont explicitement présentées.

Nom du fichier : NOx_RHBrown_20040830_R0.ict

41, 1001
Williams, Eric
Earth System Research Laboratory/NOAA
Nitric oxide and nitrogen dioxide mixing ratios from R/V Ronald H. Brown
ICARTT_NEAQS
1, 1
2004, 08, 30, 2004, 12, 25
0
Start.UTC, seconds, number_of_seconds_from_0000.UTC
9
1, 1, 1, 1, 1, 1, 1, 1, 1
-9999, -9999, -9999, -9999, -9999, -9999, -9999, -9999, -9999
Stop.UTC, seconds
Mid.UTC, seconds
DLat, deg_N
DLon, deg_E
Elev, meters
NO_ppbv, ppbv
NO_1sig, ppbv
NO2_ppbv, ppbv
NO2_1sig, ppbv
0
18
PI_CONTACT_INFO: 325 Broadway, Boulder, CO 80305; 303-497-3226; email:eric.j.williams@noaa.gov
PLATFORM: NOAA research vessel Ronald H. Brown
LOCATION: Latitude, longitude and elevation data are included in the data records
ASSOCIATED_DATA: N/A
INSTRUMENT_INFO: NO: chemiluminescence; NO2: narrow-band photolysis/chemiluminescence
DATA_INFO: All data with the exception of the location data are in ppbv. All oneminute averages contain at least 35seconds of data, otherwise missing.
UNCERTAINTY: included in the data records as variables with a _1sig suffix
ULOD_FLAG: -7777
ULOD_VALUE: N/A
LLOD_FLAG: -8888
LLOD_VALUE: N/A, N/A, N/A, N/A, N/A, 0.005, N/A, 0.025, N/A
DM_CONTACT_INFO: N/A
PROJECT_INFO: ICARTT study; 1 July-15 August 2004; Gulf of Maine and North Atlantic Ocean
STIPULATIONS_ON_USE: Use of these data requires PRIOR OK from the PI
OTHER_COMMENTS: N/A
REVISION: R0
R0: No comments for this revision.
Start.UTC, Stop.UTC, Mid.UTC, DLat, DLon, Elev, NO_ppbv, NO_1sig, NO2_ppv, NO2_1sig
43200, 43259, 43229, 41.00000, -71.00000, 15, 0.555, 0.033, 2.220, 0.291
43260, 43319, 43289, 41.01234, -71.01234, 15, 10.333, 0.522, 31.000, 0.375

2.4 Spécificités de format pour les fichiers ICARTT multidimensionnels

2.4.A Structure

Les formats de fichier ICARTT multidimensionnels sont conçus sur la base des formats de fichier standard Ames FFI=2110 et FFI=2310 ; nous recommandons l'usage de ces FFI pour l'échange de la plupart des fichiers de données multidimensionnels. Le descriptif des FFI est le suivant :

FFI 2110 ; deux variables réelles indépendantes, l'une non bornée, l'autre bornée, dont les valeurs sont entrées dans le fichier de données ; les variables primaires sont réelles ; la première variable

auxiliaire est NX(m,1) (ou variable primaire ' dimension du tableau), toutes les autres variables auxiliaires sont réelles.

FFI 2310 ; deux variables réelles indépendantes, l'une non bornée, l'autre bornée, dont le nombre des valeurs constantes d'incrément, valeur de base et incrément, est défini dans la liste des variables auxiliaires ; les variables primaires sont réelles ; les variables auxiliaires sont réelles.

Pour plus d'informations sur ces types de fichiers, veuillez consulter les documents suivants (en anglais) :

<http://www-air.larc.nasa.gov/missions/etc/Amend2110.htm>

<http://www-air.larc.nasa.gov/missions/etc/Amend2310.htm>

2.4.B. Exemples

Deux exemples de types de fichiers FFI 2110 et FFI 2310 sont présentés ci-dessous.

Exemple : FFI 2110

Nom du fichier : AR_DC8_20050203_R0.ict

```
54, 2110
PI LastName, First Name
Code 916, Goddard Space Flight Center, Greenbelt, MD 20771
AROTAL
PAVE Mission
1, 1
2005, 02, 03, 2006, 01, 18
1
Altitude[], meters, Altitude_array
UTC, XX.XXXX_hours_from_0_hours_on_flight_date
7 ;{Number of PRIMARY variables}
0.1, 0.0001, 0.1, 0.01, 0.0001, 0.1, 0.0001
-9999, -999999, -999999, -999999, -999999, -99999, -999999
TempK[], K, Temperature_array
Log10_NumDensity[], part/cc, Log10_NumDensity_array
TempK_Err[], K, Temperature_error_array
AerKlet[], Klet, Aerosol_array
Log10_O3NumDensity[], part/cc, Log10_Ozone_NumDensity_array
O3_MR[], ppb, Ozone_mixing_ratio_array
Log10_O3NumDensity_Err[], part/cc, Log10_NumDensity_error_array
11 ;{Number of AUXILIARY variable}
1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0
-9999, -9999, -9999, -9999, -9999, -9999, -9999, -9999, -9999, -9999, -9999
NumAlts, none, Number_of_altitudes_reported
Year, UT
Month, UT
Day, UT
AvgTime, xxx.x_minutes, Averaging_time_of_presented_data
Latitude, degrees
Longitude, degrees
PAlt, meters, pressure_altitude
```

GPSAlt, meters, GPS_altitude
SAT, K, Static_air_temperature
SZA, degrees
0
18
PI_CONTACT_INFO: Enter PI Address here
PLATFORM: NASA DC8
LOCATION: Lat, Lon, and Alt included in the data records
ASSOCIATED_DATA: N/A
INSTRUMENT_INFO:N/A
DATA_INFO:N/A
UNCERTAINTY: Contact PI
ULOD_FLAG: -7777
ULOD_VALUE: N/A
LLOD_FLAG: -8888
LLOD_VALUE: N/A
DM_CONTACT_INFO: Enter Data Manager Info here
PROJECT_INFO: PAVE MISSION: Jan-Feb 2005
STIPULATIONS_ON_USE: Use of these data should be done in consultation with the PI
OTHER_COMMENTS: N/A
REVISION: R0;
R0: Version 2005-0: AROTAL T & O3 Rayleigh Retrievals. Further revisions may be needed to fine-tune aerosol characterization.
UTC, NumAlts, Year, Month, Day, AvgTime, Latitude, Longitude, PAlt, GpsAlt, SAT, SZA, Altitude[], TempK[],
Log10_NumDensity[], TempK_Err[], AerKlet[], Log10_O3NumDensity[], O3_MR[], Log10_O3NumDensity_Err[]
54000, 9, 2005, 2, 3, 0, 42.308, -70.582, 6910, 6979, 242.5, 65.5
9154, -9999, -999999, -9999, -9999, 113178, 212, -999999
9304, -9999, -999999, -9999, -9999, 123353, 2250, -999999
9454, -9999, -999999, -9999, -9999, 123008, 2116, -999999
9604, -9999, -999999, -9999, -9999, 120933, 1337, -999999
9754, -9999, -999999, -9999, -9999, 119675, 1019, -999999
9904, -9999, -999999, -9999, -9999, 122655, 2061, -999999
10054, -9999, -999999, -9999, -9999, 124384, 3126, -999999
10204, -9999, -999999, -9999, -9999, 124632, 3371, -999999
10354, -9999, -999999, -9999, -9999, 121341, 1609, -999999
54001, 8, 2005, 02, 03, 0, 42.278, -70.613, 6978, 7043, 241.7, 65.5
10118, 9999, -999999, -9999, -9999, 124458, 3205, -999999
10268, -9999, -999999, -9999, -9999, 123160, 2421, -999999
10418, -9999, -999999, -9999, -9999, 121221, 1582, -999999
10568, -9999, -999999, -9999, -9999, 120950, 1523, -999999
10718, -9999, -999999, -9999, -9999, 117339, 680, -999999
10868, -9999, -999999, -9999, -9999, 122751, 2423, -999999
11018, -9999, -999999, -9999, -9999, 124230, 3491, -999999
11168, -9999, -999999, -9999, -9999, 124039, 3424, -999999

{Notez l'utilisation de facteurs d'échelle dans cet exemple.}

Exemple 2310

Nom du fichier : LIDARO3_WP3_20040830_R0.ict

46, 2310
Williams, Eric
NOAA/Earth System Research Laboratory
Ozone number density profile from WP3 aircraft LIDAR
ICARTT_ITCT
1, 1
2004, 08, 30, 2009, 09, 04

1
Geo_Alt, meters, Geometric_altitude_of_observation
UT_TIME, seconds, Elapsed_time_from_0_hours_on_day_given_by_date
1 ;{Number of PRIMARY variables}
1.0e9
-9999
O3_NumDensity[], molecules/cc, Ozone_NumDensity_Array
9 ;{Number of AUXILIARY variable}
1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0
-9999, -9999, -9999, -9999, -9999, -9999, -9999, -9999, -9999
Num_Altitudes, none, number_of_altitudes_at_current_time_mark
Geo_Alt_Begin, meters, geometric_altitude_at_which_data_begin
Alt_Increment, meters, altitude_increment_between_observations
Geo_Alt_Aircraft, meters, geometric_altitude_of_aircraft
UT_hour, hours
UT_min, minutes
UT_sec, seconds
Lon_aircraft, degrees_E
Lat_aircraft, degrees_N
0
18
PI_CONTACT_INFO: 325 Broadway, Boulder, CO 80305; 303-497-3226; eric.j.williams@noaa.gov
PLATFORM: NOAA WP3
LOCATION: Lat, Lon, and Alt data included in the data records
ASSOCIATED_DATA: N/A
INSTRUMENT_INFO: Differential absorption LIDAR. See Williams et al., BigScience, 42, p. 50-51, 2001
DATA_INFO: The units are number density (#/cc). The vertical averaging interval is 975 m at 1-7 km above the aircraft and 2025
m > 7 km above the aircraft. Horizontal averaging interval: 60 km.
UNCERTAINTY: Contact PI
ULOD_FLAG: -7777
ULOD_VALUE: N/A
LLOD_FLAG: -8888
LLOD_VALUE: N/A
DM_CONTACT_INFO: Contact PI
PROJECT_INFO: ICARTT study; 1 July-15 August 2004
STIPULATIONS_ON_USE: Use of these data requires PRIOR OK from the PI
OTHER_COMMENTS: N/A
REVISION: R0
R0: No comments for this revision.
UT_TIME, Num_Altitudes, Geo_Alt_Begin, Alt_Increment, Geo_Alt_Aircraft, UT_hour, UT_min, UT_sec, Lon_aircraft,
Lat_aircraft, O3_NumDensity[]
30335, 26, 12819, 75, 10389, 8, 25, 35, -133.24, -9.45
1340, 1519, 1660, 1779, 1868, 1939, 1973, 1992, 1989, 1955, 1934, 1897, 1817, 1721, 1619, 1514, 1434, 1343, 1258, 1203,
1140, 1088, 1037, 956, 892, 878
30336, 22, 12819, 75, 10383, 8, 26, 0, -133.22, -9.93
1351, 1523, 1658, 1774, 1860, 1930, 1962, 1974, 1966, 1932, 1909, 1877, 1803, 1706, 1600, 1493, 1407, 1310, -9999, -9999,
1094, 1045

{Notez que ce fichier utilise un facteur d'échelle (1e9) pour les données de densité car il serait encombrant d'ajouter une notation exponentielle à chaque valeur. Aussi, cet exemple a été adapté du document de la NASA et ne possède pas d'incertitude ou d'étiquettes associées aux données.}

2.5 Formats de fichiers pour des données aériennes non standard

Les données recueillies par satellite ne s'intègrent pas aisément dans le format ICARTT. Le format de données permet à chaque enregistrement de données d'être identifié par une unique étiquette de temps, uniquement lorsque les données sont rapportées de façon continue à un

intervalle de temps constant (p.ex. 1seconde). Dans le cas contraire, les temps de début et de fin doivent être rapportés et un intervalle de temps de 0 est inscrit à la ligne 8 de l'entête du fichier. Les données satellitaires sont uniques dans ce sens que, bien qu'elles soient enregistrées à un intervalle de temps constant, d'importantes lacunes dans les données peuvent subvenir. Ces absences de données peuvent être dues à des interférences avec les nuages, des changements de mode de visualisation (p.ex. observation en nadir ou au limbe) ou d'autres considérations. Étant donné l'important volume de données et la taille des fichiers associés aux observations satellitaires, il ne serait pas judicieux de combler ces lacunes dans les données par des valeurs de données manquantes. Il n'est également pas raisonnable de rapporter le temps de début et de fin car les données sont généralement collectées sur de très courtes échelles de temps (typiquement, inférieures à la seconde), si bien que la question de l'intégration du temps ne se pose pas. Au lieu de cela, les fichiers de données satellitaires rapportent un **intervalle de données** de -1 à la ligne 8 de l'entête. Cela signifie que chaque enregistrement de données s'identifie par une étiquette de temps unique, mais la chronologie réelle est discontinue. Le format ICARTT ne supporte **d'intervalle de données** de -1 pour aucune autre mesure que celles provenant des instruments satellites.

Dans certains cas, le format standard ICARTT des séries temporelles s'adapte difficilement à certaines données non standard. L'équipe de gestion des données doit considérer, au cas par cas, l'utilisation de fichiers standard communs au sein de la communauté des utilisateurs, avec l'accord préalable de l'équipe scientifique de la mission. Par exemple, de nombreux ensembles de données issus de la modélisation stockent leurs données au format NetCDF (« Network Common Data Form »), lequel est de fait standard au sein de cette communauté. Cependant, le format de données multidimensionnel défini plus haut est capable de supporter ces ensembles de données et nous laissons ce format en option. Pour certains instruments (p.ex. les lidars), les données sont disponibles en tant que fichiers d'images, habituellement sous des formats standard tels que GIF ou JPEG. Tous les logiciels permettant de lire ou de composer ces formats n'autorisent pas l'ajout d'informations additionnelles sous forme de texte (comme l'entête p.ex.). Les noms de tels fichiers doivent donc intégrer le plus d'informations possibles. Si nécessaire, l'équipe de gestion des données doit travailler en collaboration avec les chercheurs principaux responsables de ces données pour parvenir à une solution mutuellement acceptable.

2.6. Logiciels de balayage de fichier

Une suite logicielle « FScan » a été développée pour le balayage de fichiers de données et pour vérifier si les fichiers sont conformes au format standard ICARTT. La fonction de balayage effectue une vérification approfondie du fichier pour valider sa conformité ; le document est vérifié ligne par ligne, valeur par valeur et dans certains cas, lettre par lettre. Un rapport détaillé est généré, fournissant les messages d'erreur, s'il y en a, avec les lignes problématiques et les raisons de l'erreur. « FScan » offre une version en ligne ou autonome (voir les adresses web ci-dessous). De plus amples détails sur FScan sont disponibles (en anglais) à :

<http://www-air.larc.nasa.gov/missions/etc/helpFscan.html>.

Il existe deux versions disponibles pour le balayage des fichiers au format ICARTT :

1. Version en ligne : <http://www-air.larc.nasa.gov/cgi-bin/fscan>
2. Version autonome (Windows seulement) : <http://www-air.larc.nasa.gov/missions/etc/wFscan.htm>

3. Bibliographie

Références relatives au document

[1] Spécifications du format Maes de la NASA pour l'échange de données (en anglais) :

http://espoarchive.nasa.gov/archive/docs/formatspec_2_0.html

[2] Format d'archive de données NASA GTE (en anglais) :

http://www-gte.larc.nasa.gov/trace/TP_APP-E.htm

[3] Document officiel du format de données ICARTT (en anglais) :

<http://www-air.larc.nasa.gov/missions/etc/IcarttDataFormat.htm>

Références pour de l'information complémentaire :

[4] Campagne de l' « International Consortium for Atmospheric Research on Transport and Transformation » (ICARTT) (en anglais) :

<http://www.esrl.noaa.gov/csd/ICARTT/>

[5] Politique et plan de gestion des données ARCTAS (en anglais) :

http://www-air.larc.nasa.gov/missions/arctas/docs/arctas_data_plan.pdf

[6] Politique et plan de gestion des données MILAGRO (en anglais) :

http://www.eol.ucar.edu/projects/milagro/data/MILAGRO_DataPolicy.html

4. Adresse des auteurs

Ali Aknan, ali.a.aknan@nasa.gov, NASA/LaRC, MS 927, Hampton, VA 23681

Gao Chen, gao.chen@nasa.gov, NASA/LaRC, MS 483, Hampton, VA 23681

James Crawford, james.h.crawford@nasa.gov, NASA/LaRC, MS 483, Hampton, VA 23681

Eric Williams, eric.j.williams@noaa.gov, NOAA/ESRL, 325 Broadway, Boulder, CO 80305

5. Annexe A

Glossaire des acronymes

<u>Acronyme</u>	<u>Description</u>
ARCTAS	« Arctic Research of the Composition of the Arctic Research of the Composition of the Troposphere from Aircraft and Satellites »
UE	Union Européenne
GPS	« Global Positioning System », assistant de navigation
GTE	« Global Tropospheric Experiment »
Hz	Hertz
ICARTT	« International Consortium for Atmospheric Research on Transport and Transformation »
INTEX-B	« Intercontinental Chemical Transport Experiment – Phase B »
INTEX-NA	« Intercontinental Chemical Transport Experiment – North America »
ITCT	« Intercontinental Transport and Chemical Transformation »
ITOP	« Intercontinental Transport of Ozone and Precursors »
LIDAR	« Light Detection And Ranging »
MILAGRO	« Megacity Initiative :Local and Global Research Observations »
N/A	Non applicable
NASA	« National Aeronautics and Space Administration »
NEAQS	« New England Air Quality Study »
NOAA	« National Oceanic and Atmospheric Administration »
NSF	« National Science Foundation »
PI	« Principal Investigator », chercheur principal
POLARCAT	« POLar study using Aircraft, Remote sensing, surface measurements and modelling of Climate, chemistry, Aerosols and Transport »
UTC	« Universal Time Coordinated »