Defence Research and Development Canada
Recherche et développement pour la défense Canada

DRDC | RDDC
technologysciencetechnologie

A tool for expanding and exporting compact Excel databases

Standard operating procedure and application to the CSIAPS database

Sacha D. Nandlall
DRDC – Ottawa Research Centre

Defence Research and Development Canada
Reference Document
DRDC-RDDC-2017-D144
December 2017

Canada

IMPORTANT INFORMATIVE STATEMENTS

This document was reviewed for Controlled Goods by DRDC using the Schedule to the Defence Production Act.

Endorsement statement: This publication has been published by the Editorial Office of Defence Research and Development Canada, an agency of the Department of National Defence of Canada.
Inquiries can be sent to: Publications.DRDC-RDDC@drdc-rddc.gc.ca.

Abstract

The Commercial Satellite Imagery Acquisition Planning System (CSIAPS) is an application designed by Defence Research and Development Canada (DRDC). CSIAPS assists collection managers in planning acquisitions of satellite images and Automatic Identification System (AIS) ship navigational data.

The database of platform parameters used by CSIAPS is currently updated at DRDC using a shared Excel file containing multiple spreadsheets. Since many satellites and sensors share similar properties—particularly those from the same vendor or constellation—it is desirable to reduce repetitions in this Excel file in order to minimize file size and manual error. However, the CSIAPS application is designed to ingest data that is normalized and atomic, which makes it difficult to create a compact Excel file in cases where much of these data are repeated.

This document describes a tool that allows the DRDC Excel database of platform parameters to be expressed in a compact (but non-atomic) format and automatically exported to a normalized (but larger) version for ingestion into CSIAPS. The syntax for expressing data in compact form within Excel is described, followed by a standard operating procedure for exporting the compact database into a normalized form that can be imported into CSIAPS. The tool described herein could also be adapted to expand other compact Excel spreadsheets before being sent to a larger database or service.

Significance for defence and security

The tool described in this document saves time and reduces the possibility for manual error involved in keeping the CSIAPS database of satellites updated. Collection managers at the CAF as well as any partners using CSIAPS will thus obtain satellite coverage opportunities that are based on more accurate and up-to-date information. The tool described herein can also easily be adapted to other use cases in which data can be updated locally in a small, easy-to-use Excel file before being sent to a larger database or service.

Abstract

Le système de planification d'acquisition d'images de satellites commerciaux (SPAISC) est une application mise au point par Recherche et Développement pour la Défense Canada (RDDC). Le SPAISC aide les gestionnaires de collections à planifier l'acquisition d'images de satellite ainsi que des données de navigation maritime provenant du Système d'Identification Automatique (SIA).

La base de données SPAISC qui contient les paramètres de platforme est présentement mise à jour au RDDC avec un fichier Excel contenant plusieurs feuilles de calcul. Puisque beaucoup de satellites et de capteurs partagent des propriétés similaires—particulièrement ceux qui proviennent de la même compagnie ou d'une même constellation—il est désirable de réduire le nombre de répétitions dans le fichier Excel, afin de minimiser la taille du ficher ainsi que les erreurs manuelles. Toutefois, l'application SPAISC est conçue pour ingérer des données qui sont normalisées et atomiques, ce qui rend difficile la création d'un ficher Excel compact lorsque beaucoup de ces données sont répétées.

Ce document décrit un outil qui permet à la base de données Excel du RDDC d'être exprimée dans un format compact (mais non-atomique) et qui peut être exportée automatiquement dans un format normalisé (mais de plus grande taille) pour être importé dans l'application SPAISC. La syntaxe pour exprimer les données de manière compact en Excel est décrit, suivi d'une procédure pour exporter la base de données compact vers un format normalisé qui peut être importé dans CSIAPS. L'outil décrit dans ce document pourrait également être adapté pour exporter d'autres bases de données Excel compactées avant d'être envoyé vers un service ou une autre base de données plus large.

Importance pour la défense et la sécurité

L'outil décrit dans ce document réduit le temps requiert pour garder la base de données satellite du SPAISC mis à jour ; il réduit aussi la possibilité d'erreur manuelle dans cette tâche. Les gestionnaires de collections aux forces armées canadiennes, ainsi que tout partenaire utilisant la SPAISC, obtiendront donc des opportunités de couverture satellite que sont basées sur de l'information qui est plus exacte et plus à jour. L'outil décrit dans ce document pourrait aussi être adapté pour d'autres scénarios d'utilisation où des données peuvent être mises à jour localement avec un petit fichier Excel qui est relativement facile à utiliser, avant que ces données soient envoyées vers un service ou une autre base de données plus large.

Table of contents

List of figures

List of tables

Acknowledgements

The author would like to thank Michael A. Salciccioli and Jeff Secker for their assistance.

1 Introduction

The Commercial Satellite Imagery Acquisition Planning System (CSIAPS) is an application designed by Defence Research and Development Canada (DRDC) to assist collection managers in planning acquisitions of satellite images and Automatic Identification System (AIS) ship navigational data [1, 2]. CSIAPS finds satellite coverage opportunities by combining a database of satellite platforms and their key sensor parameters with orbital modelling software and a graphical user interface that lets the user specify their imaging requirements.

Keeping the database of satellite and sensor parameters updated is critical to the accuracy and effectiveness of CSIAPS. This task is currently performed at DRDC using a shared Excel file containing multiple spreadsheets. Since many satellites and sensors share similar properties—particularly those from the same vendor or constellation—it is desirable to reduce repetitions in this Excel file in order to minimize file size and manual error.

However, the CSIAPS application itself ingests data based on Comma-Separated Value (CSV) files separated by platform, and moreover expects these data to be normalized. A more detailed discussion of the topic of database normalization is available elsewhere [3, 4], but for present purposes, it is sufficient to note that an un-normalized database is typically more difficult to import, parse, and check for consistency in an automated manner. A key requirement of database normalization is atomicity, which mandates that no cell should contain multiple values. Unfortunately, when there are many repeated values or rows for which most (but not all) of the columns are identical, respecting atomicity means increasing either the size of the database or the number of tables and spreadsheets that comprise it. In both cases, this makes the database more difficult to update manually in an environment such as Excel.

This document describes a tool known as a "macro" and written in Visual Basic for Applications (VBA). The purpose of this tool is to allow the Excel database of satellite and sensor parameters to be expressed in a compact (but non-atomic) format, and then automatically export this database to a larger (but normalized) version for ingestion into CSIAPS. The syntax for expressing data in compact form within Excel is described, followed by a standard operating procedure for exporting the compact database into a normalized form that can be imported into CSIAPS.

2 Database syntax and conventions

2.1 Rows with splits

The tool described in this document allows the user to condense database rows in cases where most column values are repeated and only one or a few column values change between rows. As an example, consider the data shown in Table 1.

Table 1 : *A table with many repeated columns.*

Column A	Column B	Column C	Column D
1	foo	10.1	alpha
2	foo	10.1	alpha
3	foo	10.1	alpha
4	foo	10.1	alpha
A	bar	20.5	beta
B	bar	20.5	beta
C	bar	20.5	beta

In Table 1, the data in Columns B–D take on one set of values for rows where Column A is a number and a different set of values for rows with a letter in Column A. Although this table respects the rules of atomicity and normalization, there are several disadvantages for users manually updating this database:

- The extensive duplication increases the size of the database;
- There is an increased opportunity for manual error in cases where all values in a column must be changed for a particular group, especially if the rows are not sorted; and
- Inserting a new row involves copy-pasting all of the repeated columns, which also increases the possibility of manual error.

These issues become more significant as the size of the database grows, and can make the updating process unwieldy if tens or hundreds of rows contain mostly repeated values. In the case of CSIAPS, the database contains fifteen satellites at the time of this writing, but this number is expected to increase to hundreds or thousands in the near future. Moreover, each satellite is currently associated with approximately one hundred different sensor parameters, each of which requires one column in the database. Tracking this many satellites and sensor parameters in an accurate and timely manner is only feasible if steps are taken preemptively to reduce manual labour.

To address these issues, the tool described herein introduces the concept of a "split," which allows the columns that are not repeated to be expressed compactly. Table 2 illustrates this concept for the data in Table 1.

Table 2 : *The data in Table 1 expressed more compactly with splits.*

Column A	Column B	Column C	Column D
[1,2,3,4]	foo	10.1	alpha
[A,B,C]	bar	20.5	beta

In Table 2, Column A is now expressed succinctly with two splits: "[1,2,3,4]" and "[A,B,C]". The former expands to the first four rows in Table 1, while the latter expands to the last three rows. When the expansion is performed, the values in all of the other columns are simply repeated.

In general, a split has three components in a specific order:

1. A left square bracket "[" (if more than one left square bracket is present, the first and leftmost one is taken as the marker for the start of the split);
2. A comma-separated list of values (in this case "1,2,3,4" or "A,B,C"); and
3. A right square bracket "]" (if more than one right square bracket is present, the last and rightmost one is taken as the marker for the end of the split).

Note that the first row is assumed to contain column header information and is always copied as-is, so splits can only be used from the second row onward.

In the example shown in Tables 1 and 2, excluding the first row of column titles, using splits reduced the number of rows from seven (Table 1) to just two (Table 2). The trade-off, however, is that data expressed with splits fail to exhibit atomicity and are therefore not normalized, since atomicity requires that each table cell contain no more than one value. The tool described in this document thus allows the best of both worlds: users entering information into the database or updating it can work with the compact—but non-normalized—version containing splits (Table 2 in this example), while automated ingestion systems that use these data can import the normalized version generated by the tool, in which the splits have been automatically expanded to respect atomicity (Table 1 in this example).

The next sections describe additional conventions for using splits, as well as more powerful ways of leveraging their capability to reduce database size.

2.2 Prefixes and suffixes with splits

Any text present before the start of a split (a prefix) or after the end of a split (a suffix) is simply repeated in each row when the split is expanded. This allows some repetition to be avoided, such as in the example shown in Tables 3 and 4 below.

Table 3 : *A split with a prefix ("Room-") and a suffix ("A").*

Column A
Room-[1,2,3]A

Table 4 : *The data in Table 3 after expansion.*

Column A
Room-1A
Room-2A
Room-3A

In this example, the prefix "Room-" comes before the split and the suffix "A" follows it; thus, "Room-[1,2,3]A" expands across three rows to "Room-1A," "Room-2A," and "Room-3A." This avoids having to write "[Room-1A, Room-2A, Room-3A]," which would produce the same result but contains more repeated characters.

2.3 Whitespace within splits

Any whitespace before and after each item in the comma-separated list that describes a split is trimmed and ignored. However, whitespace between words in each list item is conserved. As an example, consider the three tables shown in Table 5.

Table 5 : *Three splits with different amounts of whitespace.*

Column A
[this,is an,example]

Column A
[this, is an, example]

Column A
[this , is an , example]

All three of these tables will expand to the same, identical result shown in Table 6. Note that the whitespace surrounding each item in the list has been removed, including any spaces around the commas that separate the list items. However, any spaces inside each component of the comma-separated list—such as, in this case, the space between the words "is" and "an"—are kept intact.

Table 6 : *The three splits in Table 5, after expansion (identical result for all three).*

Column A
this
is an
example

2.4 Multiple columns with splits

In the case where two or more columns in the same table contain splits, all possible combinations of values across these columns are enumerated. In the resulting enu-

meration, the columns furthest to the right are incremented first. As shown by the example in Tables 7 and 8, this construction is practical when all possible combinations of two or more columns have common properties.

Table 7 : *An example of multiple columns with splits.*

Column A	Column B	Column C	Column D	Column E
[foo,bar]	[1,2,3]	[A,B]	word	8.2

Table 8 : *The data in Table 7 after expansion.*

Column A	Column B	Column C	Column D	Column E
foo	1	A	word	8.2
foo	1	B	word	8.2
foo	2	A	word	8.2
foo	2	B	word	8.2
foo	3	A	word	8.2
foo	3	B	word	8.2
bar	1	A	word	8.2
bar	1	B	word	8.2
bar	2	A	word	8.2
bar	2	B	word	8.2
bar	3	A	word	8.2
bar	3	B	word	8.2

2.5 Duplicated splits

Duplicated splits are denoted by the special code of an empty set of square brackets (i.e., "[]"). Wherever this code appears, the value of the most recent split—i.e., the closest column to the left containing a split—is copied at that location. This is helpful in cases where the same index repeats in two or more columns. The examples in Tables 9 and 10 as well as Tables 11 and 12 illustrate how duplicated splits may be used.

Table 9 : *An example of duplicated splits (Columns B and C).*

Column A	Column B	Column C
[1,2,3]	System-[]	system-[]-model.mod

If none of the columns to the left of a duplicated split contain splits themselves, the duplicated split is replaced with an empty string.

Table 10 : *The data in Table 9 after expansion.*

Column A	Column B	Column C
1	System-1	system-1-model.mod
2	System-2	system-2-model.mod
3	System-3	system-3-model.mod

Table 11 : *An example of a duplicated split (Column D) with several regular splits present (Columns A, B, and E).*

Column A	Column B	Column C	Column D	Column E
[foo,bar]	Satellite-[1,2]	SAR	Model-[]	[Single, Dual]

Table 12 : *The data in Table 11 after expansion. In this case, the duplicated split in Column D is taken from Column B, since Column B is the closest column to the left of Column D that contains a split.*

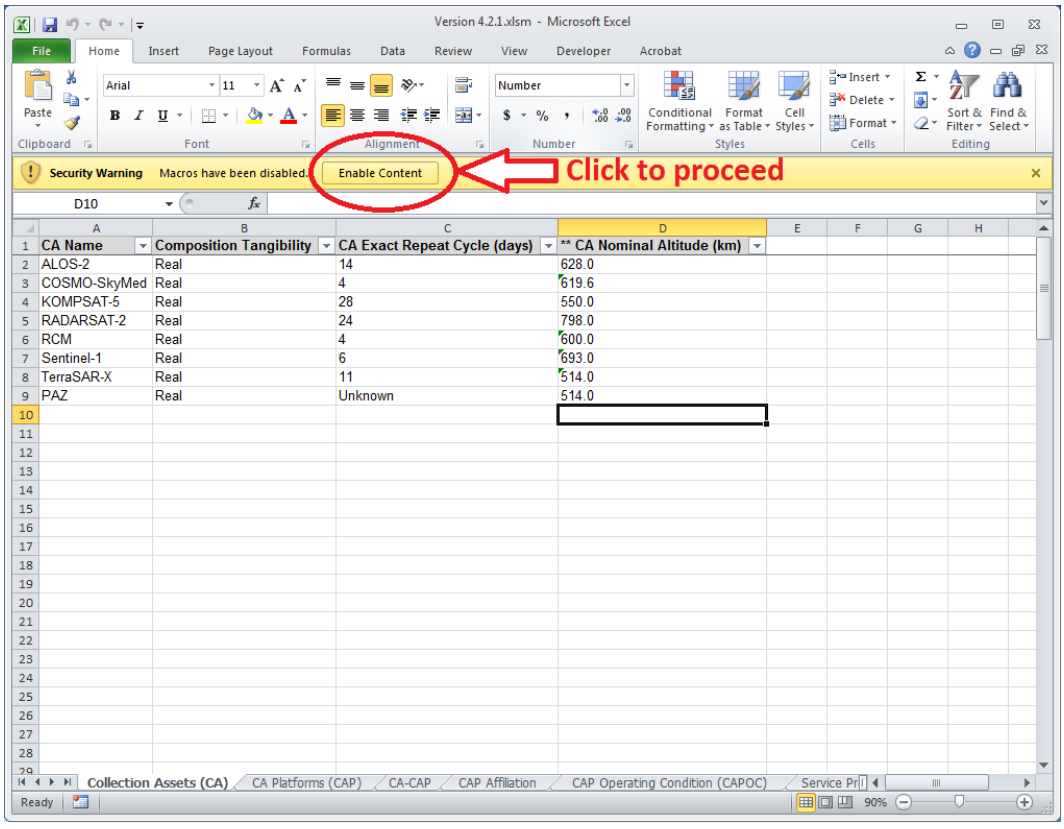
Column A	Column B	Column C	Column D	Column E
foo	Satellite-1	SAR	Model-1	Single
foo	Satellite-1	SAR	Model-1	Dual
foo	Satellite-2	SAR	Model-2	Single
foo	Satellite-2	SAR	Model-2	Dual
bar	Satellite-1	SAR	Model-1	Single
bar	Satellite-1	SAR	Model-1	Dual
bar	Satellite-2	SAR	Model-2	Single
bar	Satellite-2	SAR	Model-2	Dual

3 Procedure for expanding and exporting the database

1. In Windows Explorer, copy the macro-enabled Excel database file into a folder. Make sure write permissions are enabled for this folder, because a subfolder called "Exported Data" will be created and files will be saved there.

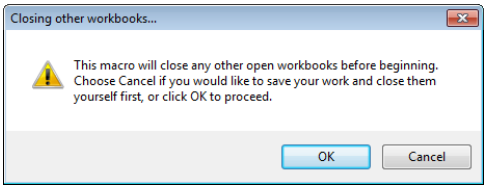
2. Open the database file in Excel and, if prompted, click the "Enable macros" button as in Figure 1. If this option is not shown, the ability to run macros may be disabled. See the relevant online support document from Microsoft for how to rectify this [5], or contact IT support if unable to change the required settings.

Figure 1 : *The button to click for enabling macros in Excel.*



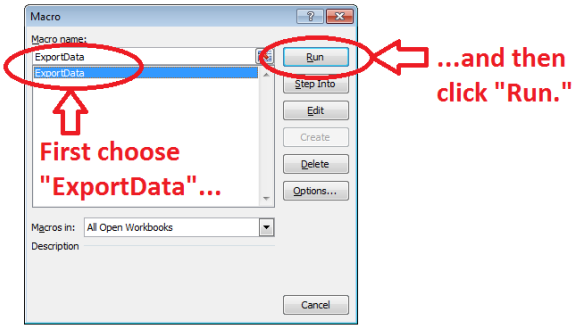
3. If any other workbooks are open in Excel, make sure to save any work done and close them. Otherwise, the prompt depicted in Figure 2 will appear after the next step. If this happens, click "OK" to proceed and close all of the other open workbooks, or click "Cancel" or to go back and save any work before this happens, and then resume the procedure at the start of the next step.

Figure 2 : *Dialog prompting the user to close any other open workbooks in Excel.*



4. Type Alt + F8 to open the macros dialog, which will look similar to Figure 3. Choose the ExportData macro in the list and then click the "Run" button.

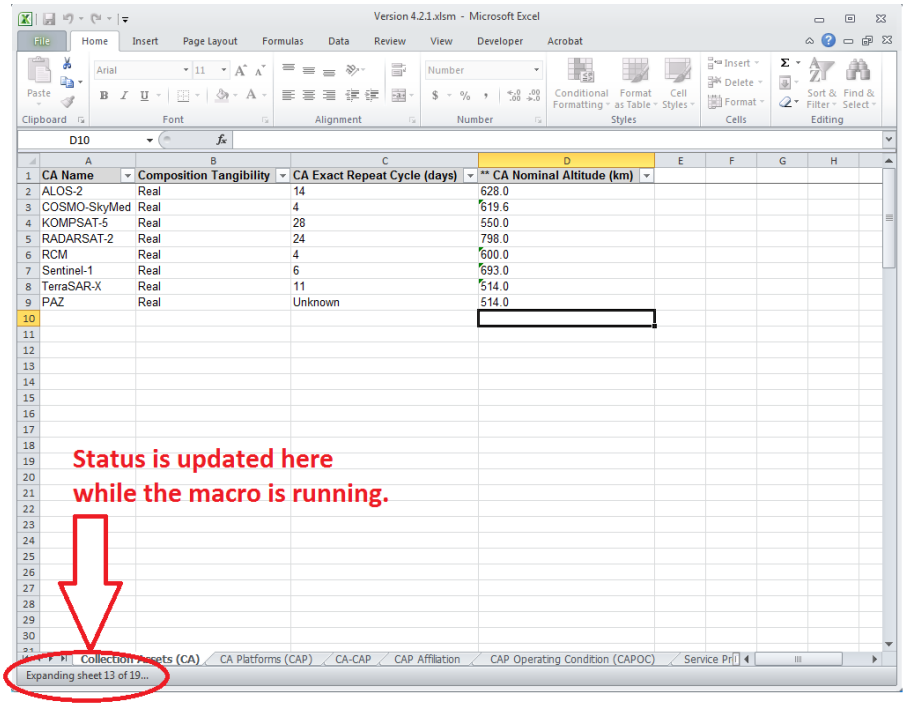
Figure 3 : *The Excel dialog for running the ExportData macro.*



5. Wait for the tool to finish executing. As it runs, Excel will be unavailable for use, but the status bar near the lower-left corner of the window will show progress updates, as illustrated in Figure 4. The main phases that require time are expanding the sheets and extracting data for each Collection Asset Platform (CAP).

6. Once the tool has finished running, switch to the Windows Explorer and go back to the folder in which the database file was saved at the first step. A subfolder named "Exported Data" should now be present. This folder will contain an Excel file of the database with all splits expanded, as well as a subfolder named "CSV" that divides the spreadsheets by CAP for ingestion into CSIAPS.

Figure 4 : *As the tool executes, progress updates are displayed in the status bar near the lower-left corner of the window.*



4 Conclusion

Using the tool described in this document, CSIAPS satellite data can be represented compactly in an Excel file and then exported automatically into a normalized database for ingestion into the CSIAPS application. Collection managers at the CAF as well as any partners using CSIAPS will thus obtain satellite coverage opportunities that are based on more accurate and up-to-date information. The tool described herein can also easily be adapted to other use cases in which data can be updated locally in a small, easy-to-use Excel file before being sent to a larger database or service.

References

[1] Secker, J. and Norris, J. (2003), Commercial Satellite Imagery Acquisition Planning System (CSIAPS): Concept Document, (DRDC-OTTAWA-TR-2003-210) Defence Research and Development Canada – Ottawa Research Centre.

[2] Robson, M., K., Cole, M. and Tse, B. (2010), CSIAPS version 2.0: Technical description, (DRDC-OTTAWA-CR-2010-198) Defence Research and Development Canada – Ottawa Research Centre.

[3] Codd, E. F. (1970), A relational model of data for large shared data banks, Communications of the ACM, 13(6), 377–387.

[4] Date, C. J. (2003), An introduction to database systems, 8 ed, New York, NY, USA: Pearson.

[5] Enable or disable macros in Office documents (online), Microsoft Inc,. https://support.office.com/en-us/article/Enable-or-disable-macros-in-Office-documents-7b4fdd2e-174f-47e2-9611-9efe4f860b12 Accessed: 2017-09-12.

DOCUMENT CONTROL DATA

(Security markings for the title, abstract and indexing annotation must be entered when the document is Classified or Protected.)

1. ORIGINATOR (The name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Centre sponsoring a contractor's report, or tasking agency, are entered in section 8.)

DRDC – Ottawa Research Centre
3701 Carling Avenue, Ottawa ON K1A 0Z4, Canada

2a. SECURITY MARKING (Overall security marking of the document, including supplemental markings if applicable.)

UNCLASSIFIED

2b. CONTROLLED GOODS

NON-CONTROLLED GOODS
DMC A
REVIEW: GCEC December 15, 2017

3. TITLE (The complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S, C or U) in parentheses after the title.)

A tool for expanding and exporting compact Excel databases

4. AUTHORS (Last name, followed by initials – ranks, titles, etc. not to be used.)

Nandlall, S. D.

5. DATE OF PUBLICATION (Month and year of publication of document.)	6a. NO. OF PAGES (Total containing information. Include Annexes, Appendices, etc.)	6b. NO. OF REFS (Total cited in document.)
December 2017	10	5

7. DESCRIPTIVE NOTES (The category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Reference Document

8. SPONSORING ACTIVITY (The name of the department project office or laboratory sponsoring the research and development – include address.)

DRDC – Ottawa Research Centre
3701 Carling Avenue, Ottawa ON K1A 0Z4, Canada

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.)	9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.)
JFD5-05EA	

10a. ORIGINATOR'S DOCUMENT NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document.)	10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.)
DRDC-RDDC-2017-D144	

11. FUTURE DISTRIBUTION OF THIS DOCUMENT WITHIN CANADA (Any limitations on further dissemination of the document, other than those imposed by security classification.)

Unlimited

12. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus. e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

CSIAPS (Commerical Satellite Imagery Acquisition Planning System), database, macro, VBA (Visual Basic for Applications), satellite, sensor, beam

13. ABSTRACT (A brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual.)

The Commercial Satellite Imagery Acquisition Planning System (CSIAPS) is an application designed by Defence Research and Development Canada (DRDC). CSIAPS assists collection managers in planning acquisitions of satellite images and Automatic Identification System (AIS) ship navigational data.

The database of platform parameters used by CSIAPS is currently updated at DRDC using a shared Excel file containing multiple spreadsheets. Since many satellites and sensors share similar properties—particularly those from the same vendor or constellation—it is desirable to reduce repetitions in this Excel file in order to minimize file size and manual error. However, the CSIAPS application is designed to ingest data that is normalized and atomic, which makes it difficult to create a compact Excel file in cases where much of these data are repeated.

This document describes a tool that allows the DRDC Excel database of platform parameters to be expressed in a compact (but non-atomic) format and automatically exported to a normalized (but larger) version for ingestion into CSIAPS. The syntax for expressing data in compact form within Excel is described, followed by a standard operating procedure for exporting the compact database into a normalized form that can be imported into CSIAPS. The tool described herein could also be adapted to expand other compact Excel spreadsheets before being sent to a larger database or service.

RÉSUMÉ (Résumé succinct du document. Le résumé peut paraître ailleurs dans le corps du document. Il est éminemment souhaitable que le résumé d'un document classifié soit sans classification. Chaque paragraphe du résumé doit commencer par une indication de la classification de sécurité des renseignements qu'il contient [sauf si tout le document est sans classification]; utiliser les lettres voulues: S, C ou U. Il n'est pas nécessaire de mettre ici le résumé dans les deux langues officielles, sauf si le document est bilingue.)

Le système de planification d'acquisition d'images de satellites commerciaux (SPAISC) est une application mise au point par Recherche et Développement pour la Défense Canada (RDDC). Le SPAISC aide les gestionnaires de collections à planifier l'acquisition d'images de satellite ainsi que des données de navigation maritime provenant du Système d'Identification Automatique (SIA).

La base de données SPAISC qui contient les paramètres de platforme est présentement mise à jour au RDDC avec un fichier Excel contenant plusieurs feuilles de calcul. Puisque beaucoup de satellites et de capteurs partagent des propriétés similaires—particulièrement ceux qui proviennent de la même compagnie ou d'une même constellation—il est désirable de réduire le nombre de répétitions dans le fichier Excel, afin de minimiser la taille du ficher ainsi que les erreurs manuelles. Toutefois, l'application SPAISC est conçue pour ingérer des données qui sont normalisées et atomiques, ce qui rend difficile la création d'un ficher Excel compact lorsque beaucoup de ces données sont répétées.

Ce document décrit un outil qui permet à la base de données Excel du RDDC d'être exprimée dans un format compact (mais non-atomique) et qui peut être exportée automatiquement dans un format normalisé (mais de plus grande taille) pour être importé dans l'application SPAISC. La syntaxe pour exprimer les données de manière compact en Excel est décrit, suivi d'une procédure pour exporter la base de données compact vers un format normalisé qui peut être importé dans CSIAPS. L'outil décrit dans ce document pourrait également être adapté pour exporter d'autres bases de données Excel compactées avant d'être envoyé vers un service ou une autre base de données plus large.