

Review of ICARTT file format implementation and operational suitability

NASA's Earth Science Data Systems Standards Process Group (SPG) is considering the International Consortium for Atmospheric Research on Transport and Transformation (ICARTT) file format for adoption as a community standard. The ICARTT file format was developed to fulfill the data management needs for the ICARTT campaign in 2004. This file format is text-based and composed of a header section (metadata) with critical data description information (e.g., data source, uncertainties, contact information, and brief overview of measurement technique), and a data section. Although it was primarily designed for airborne data, the ICARTT format proved to be practical for other mobile and ground-based studies and various data types. The ICARTT file format has since been widely accepted in the airborne field study community and used in recent major airborne studies sponsored by NASA, NSF, NOAA and international partners.

You are invited to review this Requests For Comment (RFC) in the context of your **implementation experience** with this data format specification and its **suitability for operational use**. You only need to answer questions applicable to you. Please send your completed review to:

spg-rfc-019@lists.nasa.gov.

Implementation Experience questions:

1. (*Your background*) Describe in a sentence or two your overall implementation experience related to the proposed specification. (*e.g., specification implementer, tools developer, data provider, scientific analyst, science user, etc.*) Have you directly implemented the ICARTT format specification? Did you use a pre-existing software package, and if so, what did you use?
I'm the Data Manager so I collect and check ICARTT files from scientists. I have also written tools for scientists to create and read ICARTT files.
2. (*Completeness*) Does the specification provide all the detail you need to implement it in software? (*e.g., to read or write a data file; to implement the specification, a profile or extension; or develop a tool such as a format translator*) If not, describe what is missing in the specification. Yes.
3. (*Accuracy*) Do any parts of the specification contain inaccuracies, or internal inconsistencies? If so, please provide details. Page 13 of ES-DS-RFC-019-v0.01.pdf lists the column titles in bold as an example of the last line of the header, but those titles should be comma delimited.
4. (*Clarity*) Is any part of the specification ambiguous, or poorly explained? If so, please provide details. Is there a specific file type that should be used? Mac (CR) vs PC (CRLF) vs Unix (LF) line endings? I have run into cases where a person on a PC couldn't read the file header since the file was saved with Mac line endings.

I'm also not quite clear on the possible uses of the file extension which is defined as the file type descriptor, but then the document ESDS-RFC-019-v0.01.pdf later says this is an extension of 2-4 characters to identify the file type. If it is an ICARTT file, what other extensions could there be other than .ict?

5. (*Usefulness*) How well does this specification meet your information sharing needs? (*e.g., does it work well with the data types and data manipulations in your application? Does it properly represent your datasets? What are the pros and cons of this data format?*) I think it does work well overall.
6. (*Implementation*) What implementation challenges does the proposed standard present? Please provide details, if any.

Operational Suitability questions:

7. Do you currently use or plan to use the ICARTT format in a production setting? Do you plan to distribute data in this format to science collaborators and other researchers? Yes, and yes.
8. Why do you choose to use the ICARTT format over other data formats for your applications? It is easy for a variety of users to use on different platforms, and with different analysis software.
9. Does the ICARTT file format meet your requirements for storing and accessing data? *Yes*
10. Have you or your users encountered any difficulty when using some of the data in the ICARTT format? If you have, please provide a brief description of your experience. Problems with line endings. See 4 above.
11. What operational challenges or limitations does ICARTT present? Please provide details. The ICARTT file isn't as quick to load as a binary file, but is probably worth the trade-off in ease of use for different types of computers.
12. What benefits does the ICARTT file format present? Are there any drawbacks to using this file format? (*e.g., Does it offer the flexibility you want to package the data types in your applications? Does it facilitate interdisciplinary studies?*) Benefits include, ease of use for different types of computers, and having the meta-data included in the same file as the data.
13. How much data do/will you provide or archive in the ICARTT format? (*Number of distinct data sets, total data volume, number of files.*) Currently we have about 3.5GB and 2100 files in ICARTT format from past field missions. For the upcoming CalNex project in 2010, I would expect another 1 GB or so.
14. How many users do you have or expect to have for data in the ICARTT format, and what is your expected user community? About 100 (?) users in the

tropospheric chemistry community, which includes NOAA, NASA, and several universities.

15. (*User comments*) Any additional comments, observations or criticisms of the ICARTT format and the RFC can be provided here.