|  |  |  |
| --- | --- | --- |
| WORLD METEOROLOGICAL ORGANIZATION  COMMISSION FOR BASIC SYSTEMS  -----------------------------  FOURTH MEETING OF  INTER-PROGRAMME EXPERT TEAM ON DATA REPRESENTATION MAINTENANCE AND MONITORING  GENEVA, SWITZERLAND, 30 MAY - 3 JUNE 2016 |  | IPET-DRMM-IV / Doc. 3.3 (1)  (17. 5. 2016)  -------------------------  ITEM 3.3  ENGLISH ONLY |

BUFR AND CREX

**Proposed feature selection for BUFR Edition 5**

*Submitted by* *Yves PELLETIER (Canada)*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Summary and Purpose of Document**

This document and the spreadsheet in annex summarize the requirements gathered for BUFR Edition 5. A triage of the requirements is proposed, as well as ways forward for the selected requirements.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**ACTION PROPOSED**

The meeting is requested to consider the documents and proposals, and issue recommendations regarding their implementation in BUFR Edition 5.

**ANNEXES:**

1. See accompanying electronic spreadsheet

**Context**

The IPET-DRMM III meeting in Beijing discussed BUFR Edition 5 and how to proceed with the list of requirements that has been gathered in recent years. One important issue was whether to include the following requirement, as per recommendations of the joint IPET-DRC and IPET-MDI meeting in Exeter (2012):

“A consistent approach to the handling of generalized coordinates and polymorphism [is required]. The regulations for BUFR Edition 5 must stipulate that all Table D sequences used must be opened and closed by the occurrences of the same Generalized Coordinate and/or Significance Qualifier.”  (We will refer to this henceforth as “scoping”).

While highly desirable in principle, the above would require the rewriting of many, if not most, current Table D sequences. It may also not be possible to apply scoping to all combinations of descriptors allowable by the current Data Syntax regulations. Thus, attempts to resolve the ripple effects of scoping could require the review and modification of certain existing regulations and practices in the use of the BUFR Data Syntax. In light of this, and of the short time left to propose BUFR Edition 5 for the upcoming CBS, the guidance resulting from IPET-DRMM III discussions was to focus on feature requests that would provide improvements with lesser impacts.  This is what we will now endeavour to do.

**Documentation**

An Excel table has been in existence for the past few years which documents the feature requests that have been received over time for BUFR Edition 5. The latest version of this file is attached herewith. This new version contains three sheets. The first sheet contains the proposed reduced set of Feature Requests that we propose for implementation. The second sheet contains the full list of Feature Requests before triage; and the third one documents the impact scale that was used for the purpose of triage.

In the full list, the Feature Requests are sorted in accordance to the following themes:

1.       Metadata interoperability

2.       Other functional enhancements to the code form itself

3.       Procedural enhancements, usually meaning improvement requests to the management infrastructure surrounding the Tables

In addition to this classification by theme, the proposals were triaged in accordance with an impact scale. The impact is quantified as nil, light, medium or major, as described in more detail within the Excel file, in the sheet entitled “Impact scale”.

Finally, some triage decisions were based on judgment.  In these cases, the considerations were mainly of cost/benefit, and whether a sufficiently clear and compelling use case existed according to our best judgment and knowledge, given the context described above.

Following this process, we retained those proposals from the Joint Meeting in Exeter that remained achievable given those restrictions. Regretfully, we felt we had to rule out some of the most meaningful proposals due to the complexity and ramifications of their implementation in light of the context discussed above.

**Software implementation, testing and eventual deployment**

It is proposed to achieve the software implementation of BUFR Edition 5 by means of relatively incremental changes to the existing source code of existing encoder/decoders for BUFR Edition 4.

There is currently no concrete software implementation of the proposed BUFR Edition 5. It is proposed that interested Centres should devise a development plan that would include the implementation, testing and validation of the proposed new features. The eventual operational migration to BUFR Edition 5 is envisioned as a gradual process, driven in part by the introduction of new data products. It is recommended that BUFR Edition 4 remain in operational use for a significant amount of time in the future.

**Summary of proposed changes for BUFR Ed. 5**

What follows is a list of the proposals that were selected, with associated proposals for further discussion. For full details, please see the spreadsheet in annex to this document.

|  |  |
| --- | --- |
| **Proposed Requirement** | **Proposed for further discussion** |
| The development of a Logical Data Model (Application Schema) should be the initial activity in developing a new data exchange requirement. | It is proposed to issue a recommendation encouraging the use of LDM in developing new data exchange requirements. Even in the eventuality of this becoming mandatory, this is an overarching statement that should not be part of the BUFR regulations as such. A suitable location should be found for this recommendation in the WMO Technical Regulations. |
| Enforced use of templates in product development | It is proposed to make the use of templates for new BUFR data products mandatory in the case where they are intended for exchange in the WMO framework. This is an overarching requirement that should not be part of the BUFR regulations as such. Therefore, a suitable location should be found to state this requirement in the WMO Technical Regulations. |
| Short and long names for Table B & D entries | It is proposed to add a "long name" column to Table B entries, and make the use of "short name" and "long name" mandatory in naming Table D sequences. Thought should be given to mechanisms ensuring backward compatibility of the Tables for the encoding and decoding of BUFR Edition 4. |
| Representation of values over extremely large dynamic range | A workable solution exists in BUFR Edition 4, and further debate on this matter would be unlikely to be fruitful. |
| Explicit separation of data types from units | This is relatively easy to implement in the tables. Some debate is required to hash out whether the pros outweigh the cons. |
| Multilingual support for character strings | The specifics of the implementation are yet to be determined, as well as whether the pros outweigh the cons. But if this is deemed desirable, the current window of opportunity would seem like a good one. |
| Delayed sequence definition | 1. Create new type of descriptor with F=4 (this is made possible with the proposal immediately below).  2. Use a delayed mechanism similar to F=1 to indicate the number of descriptors to be retrieved from section 4 (in a similar manner to the existing delayed replication of descriptors).   3. The data associated with the delayed sequence would follow immediately in Section 4.  Whether there is a strong use case for this is open for discussion. One possibly compelling example is that it could replace the currently existing Data Present Bitmap mechanism in a manner that could seem better integrated with the Data Syntax as a whole. |
| Increase name-space size for descriptor classes and descriptor numbers (XX and YYY in F-XX-YYY) | Increase descriptor size to 24 bits.   F-X-Y structure to be extended to 3-9-12 bits. Maintain numerical backward compatibility. That is to say: the meaning and interpretation of current Table B and D descriptors as per the numerical value of X and Y are unchanged. Table C descriptors need to be re-implemented in accordance with the new descriptor size. Maintain the current numerical ranges for local descriptors, while considering extending it by a reasonable extent.  The extension of F to 3 bits opens the possibility of new types of descriptors in BUFR edition 5 or beyond. |
| Allow for the description of the object of statistical (or other) operations when multiple possibilities exist. Example use case: specify which fields are the subject of correlation calculations in radar data. | Two possible options:  1. Add the required functionality as another variant of the Data Present Indicator (DPI) mechanism. This may be possible in BUFR Edition 4 without going to another edition if no new Table C descriptor is required.  2. Use the proposed Delayed Sequence as a means to indicate the fields which as subject to the correlation (or other operation). This would follow a similar pattern of operation as the DPB but without the need for at Table C operator. |
| Use industry standard bit order | Reverse the current order of bits ordained in note (2) of the section entitled "SPECIFICATIONS OF OCTET CONTENTS". |
| Versioning and life-cycle of Table B and D entries | For the effective application of versioning information, two conditions have to be met: messages have to be encoded with accurate version information (of Master Table Version, etc). AND the decoder and its downstream processes have to make proper use of the information. For instance, in cases where the bit width of some Table B descriptors changed from one version to the next, having the correct version number is essential to fully decode the BUFR message.  There are Code Form regulations specifying how to convey the requisite information, but each Centre is ultimately responsible for its own actions in this regard. QA/QC processes, whether the Centre's own or in the WIGOS framework, should be concerned with a capacity for identifying and reporting various encoding issues, including issues of versioning. Thus, regulations concerning the application of versioning made possible in the BUFR Code Form would perhaps carry more weight if they were referred to in the Technical Regulations governing the production and transmission of messages.  With regard to the life-cycle of entries in Tables B/D, or more concretely a clean-up of flawed or obsolete entries, this is beyond the scope given to BUFR Ed. 5 by the discussions at the Beijing meeting. |
| A new methodology is needed in the new editions of BUFR and GRIB to ensure that decoders are always using the exact same table information as encoders. | A mechanism for real-time authoritative table validation would alleviate or perhaps solve this issue. This could be developed outside the direct scope of BUFR Edition 5, since it does not necessarily have to involve a change in the encoding /decoding process. |