

Universal Multiple-Octet Coded Character Set  
UCS

ISO/IEC JTC1/SC2/WG2/IRG N1906

Date: 2012-11-09

Source:	Japan
Title:	Consideration on expanding definition of evidence to accept online databases (Input to the revision of IRG PnP)
Meeting:	IRG #39
Status :	
Actions required	For Discussion
Distribution:	IRG
Medium :	Electronic
Page:	
References:	N1813ROK_Feedback, N1882

With encouragement by ROK, Japan proposes to revise the definition of "evidence".

## 0. Background

When proposing ideographs to be encoded into ISO/IEC10646, it is requested to submit "scanned image" of printed document as evidence to IRG. The original purpose of evidence is to confirm the correctness and usage of the proposed ideographs. However, it is known widely that there are ideographs that cannot be found in the dictionary or any other printed documents, but such ideographs sometimes indispensable for the specific purpose.

This document proposes to modify the definition of evidence in IRG PnP so that it matches the practical requirement.

## 1. Proposal

"Evidence" is explained in the current version of PnP as below:

2.2.4 a. Supporting Evidence: Evidence should be supplied to support the proposed glyph shape and the usage and context with pronunciations, meanings, etc., to convince the IRG that it is actually used or non-cognate with other similar ideographs. Evidence for each character must be supplied as scanned images. The provision of evidence on character usage including those for personal names should not be exempted. A declaration for character use without accompanying evidence is not acceptable. Considering privacy issues, the IRG has suggested some compromised provision. Details are given in Annex G Part 3.

Japan proposes to replace it with:

2.2.4 a. Supporting Evidence: Evidence should be supplied to support the proposed glyph shape and the usage and context with pronunciations, meanings, etc., to convince the IRG that it is actually used or non-cognate with other similar ideographs. Evidence for each character must be supplied as scanned images of printed materials or authentic online databases. The provision of evidence on character usage including those for personal names should not be exempted. A declaration for character use without accompanying evidence is not acceptable. Considering privacy issues, the IRG has suggested some compromised provision. Details are given in Annex G Part 3.

Adding to above, "authentic" should be defined clearly. Japan suggests "authentic" should mean:

- (1) Open to public
  - Anyone can access to the information via open method (e.g. internet) without special procedure (i.e. database is not limited to a closed community, etc.)
  - The minimum information (e.g. operation of the database) is offered in English. (contents can be offered in a local language, the same situation as printed documents)
- (2) Traceable
  - All modifications and reasons are recorded and can be referred at anytime.
- (3) Developed and operated by authorized committee (e.g. government organization, etc.)
- (4) Anything else ...

For the above items, Japan expects IRG editorial group will discuss and revise PnP document regarding concluded terms and conditions. When submitter is trying to use online database as evidence under the concluded conditions in the future, he/she should explain at IRG for those items.

## 2. Rationale

- (1) ISO/IEC 10646 already has encoded more than 70k CJK Unified Ideographs. Commonly used ideographs and most ideographs in the major dictionaries (including IRG dictionaries) are already encoded.
- (2) Ideographs requested to be added in the future (including CJK F) may be the following.
  - a) Ideographs for proper names, such as person's names or place name, and use of them are required by the regulation or law of government or any public organization. Some of these may be variations of ideographs that are not unified by the unification rule.
  - b) Ideographs found in academic archives. Such ideographs are not be used in general, but are used by researchers for academic purpose.
- (3) For the case of (2) a), the needs of standardize ideographs is very high for the purpose of digitization of procedure or development of the system. Specifications are often not

distributed in printed documents, is published in electronic and distributed on the internet. Such document will not meet the requirements of the "printed image", however, the documents are developed by an authorized organization, and definitely that assumes the implementation for the public service. It is seriously problematic if international standard cannot be adopted to develop public service (PUA or any other arbitrary negotiation is out of scope for this purpose).

(4) Databases that have the same feature of (2) a) are practically used in the IRG's editorial work as shown below.

- 全字庫 <http://www.cns.11643.gov.tw/>
- 香港增補字符集 (HKSCS, Hong Kong Supplementary Character Set) [http://www.ogcio.gov.hk/en/business/tech\\_promotion/ccli/hkscs/](http://www.ogcio.gov.hk/en/business/tech_promotion/ccli/hkscs/)
- Macao Supplementary Character Set

Consequently Japan expects this proposal does not affect in any way to the current policy of IRG work.

### 3. Consideration of using personal information as evidence

For some ideographs which have the same feature mentioned above, IRG accepted personal information as evidences in the past as a compromise and set considerations of handling such information as PnP AnnexG.3. In the practical operation submitter only projected on the screen to check the documents in just a short period at editorial meeting, and sharing documents is strictly prohibited because of the privacy concern. Because of this, anyone cannot verify that the document is genuine or not, and people out of IRG editorial group cannot see the document later to confirm. Such information does not satisfy the IRG's criteria for evidence and this may cause degrade the quality of CJK Unified Ideographs as a result. Accept of such materials should be avoided in the future.

To avoid this practice, Japan proposes one more thing as below.

Replace 2.2.4 a with:

2.2.4 a. Supporting Evidence: Evidence should be supplied to support the proposed glyph shape and the usage and context with pronunciations, meanings, etc., to convince the IRG that it is actually used or non-cognate with other similar ideographs. Evidence for each character must be supplied as scanned images of printed materials or authentic online databases. The provision of evidence on character usage including those for personal names should not be exempted. A declaration for character use without accompanying evidence is not acceptable. ~~Considering privacy issues, the IRG has suggested some compromised provision.~~

~~Details are given in Annex G Part 3.~~

### Delete Annex G Part 3

#### G.3. Handling of Data with Privacy Concerns

The IRG understands that the current privacy laws and practices in different Countries and Regions can make the submission of complete records as evidence related to personal information difficult. As a compromise, the IRG suggests member bodies to provide evidence in such a way that it would not reveal complete personal/internal information. However, the character information itself must be shown in the supplied evidence. In other words, partial document images should be supplied with certain sensitive information blocked.

As different departments/organizations may have different types of documents, the IRG suggests that, for each type of document, a submitter provides a sample document with private information deleted. A good example is the original Basic Certificate of Family Relation Register in Korea as seen in Fig. G1. The evidence can be submitted as partial data in the form shown in Fig. G2.

Japan believes that special ideographs for personal identities that are already included in the standard based on the above handling could have been handled with the online database evidences.

(End of Document)