ISO/IEC JTC1/SC2/WG2 N4942

Title: **Comments on Shuishu in PDAM2.2 text**

Source: **suzuki toshiya**

Document Type: individual contribution

Date: **2018/04/02**

Abstract

Shuishu script in PDAM2.2 code chart has several points to be resolved before the standardization; 2 major points are described in this document, and the possible solutions are proposed. The discussion for future standardization is needed.

1. The relationship between methodology and purpose is unclear.

1.1. Request of the clarification of the purpose.

According to N4638 [1], the initial submission, there is a paragraph; *Shuishu is strictly private. Every Shui family possesses its own version, transferring it from generation to generation. Generally, outsiders are prohibited from seeing it. Moreover, Shuishu masters usually distort the shape of the symbols to increase the difficulty of decipherment, which diversifies Shuishu characters*. It makes difficult to understand the purpose of this standardization[1] If Suishu masters are trying to improve the security or the workflow by digital technology, the proposal of the standardization is very straightforward solution. But we cannot find such movement in the materials acompanied to the submission.

In following, I assume the purpose of the standardization is the preservation of Shui culture, as endangered writing system. Or, it is an attempt to establish new writing system for daily spoken Shui language, based on Shishu? They are quite different. Although I understand some people think speedy allocation of the codepoints in ISO/IEC 10646 is far important than the clarification of the purpose how the codepoints would be used, I strongly suggest to clarify the purpose of the standardization. It is important to indicate for the users, what is done, what is future issue.

1.2. Why so strong glyph normalization? Does it serve the preservation of endangered writing system?

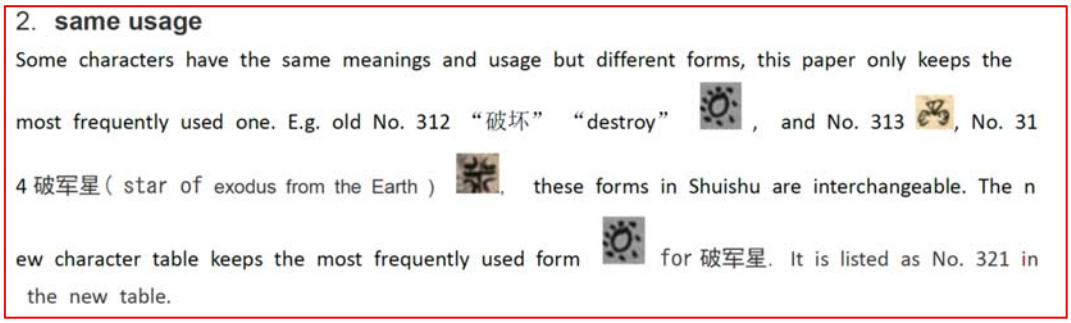
If the purpose of this standardization is the preservation of the endangered writing system, why the submitters try to normalize the glyphs in the existing documents? Of course, there would be little requirement to have multiple codepoints with same sematics and subtle shape difference. But why the glyphs with significant shape difference should be unified? Even in CJK Unified Ideograph (which is not endangered writing system, and now it is the dominant part of ISO/IEC 10646), sometimes the significantly different glyphs with almost same semantics are coded separately (e.g. 一 and 弋 are coded separately). Although it is questionable whether we should keep the separate encoding of the semantically interchangeable CJK Unified Ideographs in future, there is a variation selector technology to distinguish them.

If the normalization of Shuishu is just for the definition of the initial compact set for Shuishu, to keep from the over-disunification troubles (often found in CJK Unified Ideograph history), that's ok, but it should be stated how the

[1] This statement is already conflicting with the answers to the submission form section C (technical questions) Q4 : "the context of use for the proposed characters (type of use; common or rare), plus reference". The answer in the form is "yes, it is widely used among Shui people".

variants with significant shape difference should be handled, even if the normalization criteria is classified as the future issue. Without the clear policy about it, the researchers of Shui scripts would confront with the difficulty “oh, this glyph on this document is not found in ISO/IEC 10646, how should I digitize this text?”. This situation is questionable whether the incorporation of Shuishu character into ISO/IEC 10646 helps the study of Shuishu, or becomes yet another barrier against that.

The typical example would be the representative glyph for 破軍星, explained in WG2 N4758 [3], p.9. The glyphic difference among No. 312, No. 313 and No.314 are obviously significant.



2. **same usage**

Some characters have the same meanings and usage but different forms, this paper only keeps the most frequently used one. E.g. old No. 312 “破坏” “destroy” , and No. 313 , No. 314 破军星（star of exodus from the Earth）, these forms in Shuishu are interchangeable. The new character table keeps the most frequently used form for 破军星. It is listed as No. 321 in the new table.

Example of glyph variants (N4758 [3])

If Shuishu masters do not want to encode them separately, and do not want the external researchers to create the digital texts assigning PUA codepoints for the glyphs no. 313, 314, it should be clarified. Of course, their requests (if there is such) should be respected, but still the introduction of the variation sequence should be considered.

1.3. Questionable selection of representative glyph

The section “(1) Reasons for deletion - 3. variants” in WG2 N4758 [3] p.9 tells as if the most representative glyphs were chosen by the most frequent character of the checked materials. But there are a few questions;

A) the coverage is sufficient to choose the single representative glyph?

（1）Source texts

The source texts are 17 important Shuishu texts from 8 volumes.
Source texts list (Shuishu texts materials) for this proposal (8 volumes and 17 books). Each source text includes an interlinearized translation are published or in the process of publication.

Source Book No.	shortened form for the source texts	Source Book name (Chinese)
S1	Tsinghua Shuishu Text-10 Books	S1 《<清华大学馆藏水书文献 10 本>解读》
S2	Book of Golden Day	S2 《渤金-纪日卷》
S3	Gold and Silver	S3 《金银卷》
S4	Lu-Dao	S4 《<陆道根原>解读》
S5	Cycle of Sixty Years	S5 《<六十甲子时辰>解读》
S6	Tanju (Nine Stars)	S6 《贪巨九星歌本》
S7	Sixty Dragon	S7 《六十龙备要》
S8	Jixing (Lucky Star)	S8 《吉星》

Please see appendix 5（1）List of Shuishu characters source texts and book covers.

Source references for Shuishu submissions (N4758 [3])

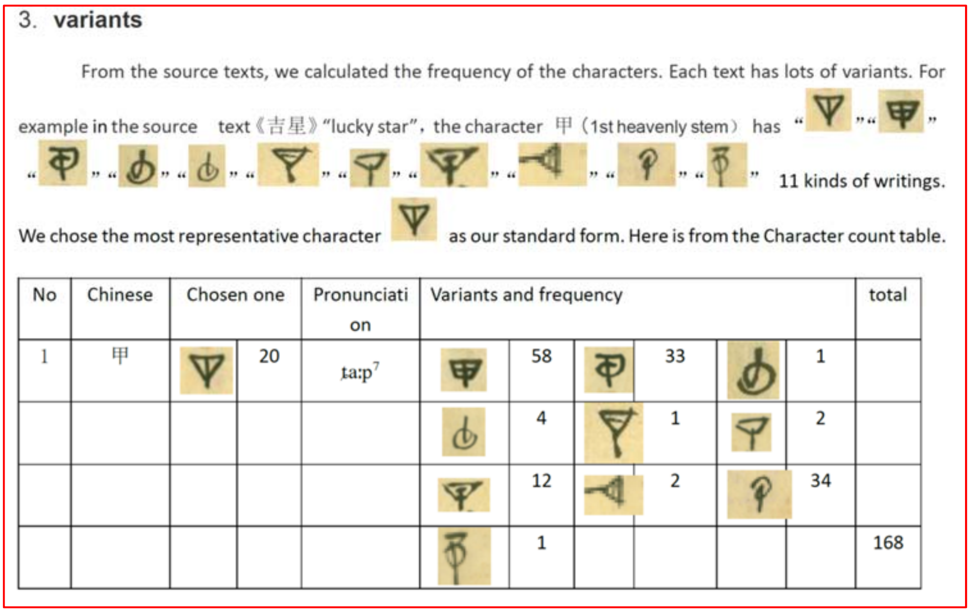
If the policy for Shuishu encoding is “no variants should be coded separately, and no variation selectors should be considered”, the representative glyphs should be chosen very carefully. According 2-(1) “Source texts” in it, 17 books (maybe S1 includes 10 books, S2-S8 include 1 book per 1 reference) are considered.

Is this sufficiently large part of Shuishu materials? Considering that “中國水書” [10] consists from 160 volumes, there is a concern whether the representative glyphs chosen by the survey for 17 books is sufficiently stable. I am afraid that some glyphs (which now the submitter dropped as less frequently used in 17 books) would be found the most frequently used glyphs in later survey. If their glyphic differences are subtle, it would be possible to update the code chart. But if their glyphs differences are significant, how to do that? The representative glyph of the code chart should be replaced by new one? Or, should be coded separately? The policy would be important for the researchers working with the books out of 17 books used for the submission.

B) unclear process how the representative glyphs were choesen

The subsection “3-(1)-3 variants” is difficult to understand how the “representative” glyph of “甲” in 水書, “ ”, was chosen. The explanation mentions about the frequency, but the chosen glyph is not the most frequently used. It is the 4th one (frequency 20). Why the 1st one “ ” (frequency 58), the 2nd one “ ” (frequency 34) and “ ” (frequency 33) are not chosen?

- If 1st one was regarded as “this is Hanzi, not Shuishu”, why the characters like (四), (七), (十) are included in the submission?
- If 2nd and 3rd ones are regarded as problematic (e.g. too cursive), why the representative glyph for “申”, , is included in the submission? The procedure how the representative glyphs are chosen, should be more clarified to understand the stability (if the submitters are against the encoding of the variants).



3. **variants**

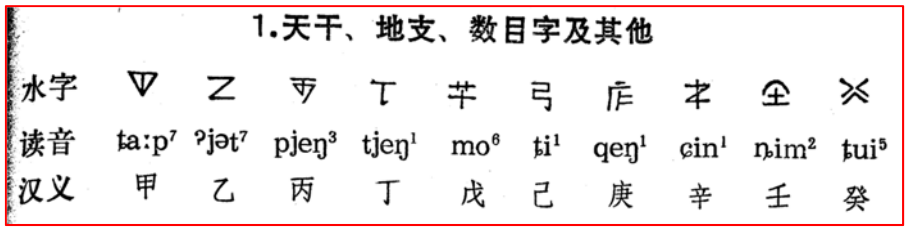
From the source texts, we calculated the frequency of the characters. Each text has lots of variants. For example in the source text《吉星》“lucky star”, the character 甲（1st heavenly stem） has “ ”“ ” “ ” “ ” “ ” “ ” “ ” “ ” “ ” “ ” “ ” 11 kinds of writings.

We chose the most representative character as our standard form. Here is from the Character count table.

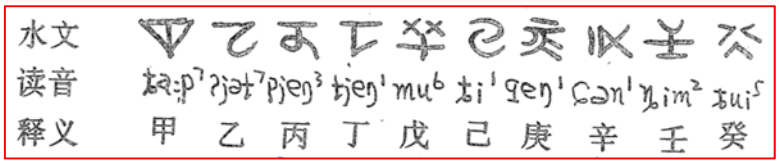
No	Chinese	Chosen one		Pronunciation	Variants and frequency						total
1	甲		20	ʈa:p7		58		33		1	
						4		1		2	
						12		2		34	
						1					168

The frequencies for the glyphs corresponding to Hanzi “甲” (N4758 [3])

I emphasize that I am not saying "  " should be chosen as the representative glyphs. It is easy to find the past researches dealing "  " as the representative glyphs for 甲, like, "水語簡志" [8] or "水書" [9].



"甲" in 『水語簡志』[8]

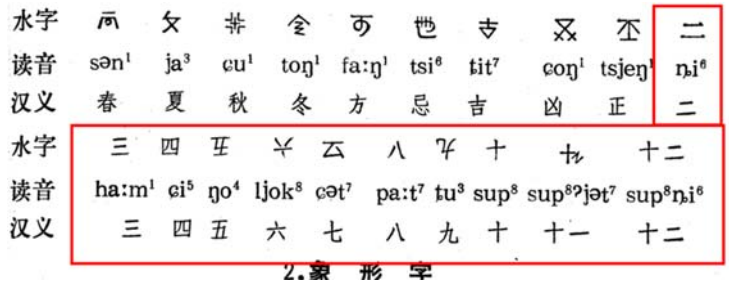


"甲" in 『水書』[9]

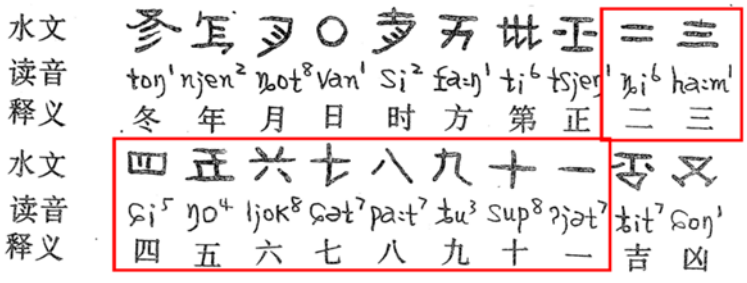
However, some glyphs in the proposals are different from the past researches. We can find the numeral 2 and 3 are rotated in the comparison with the past researches. Such differences have big impact to set the codepoint in the proposal, because the proposed radical systems have the radical for vertical stroke and horizontal stroke. The clarification of the rule to choose the representative glyphs is quite important.

1	
2	
3	
4	
5	
6	
7	
8	
9	
10	

numerals 1-10 in N4758



『水語簡志』[8]



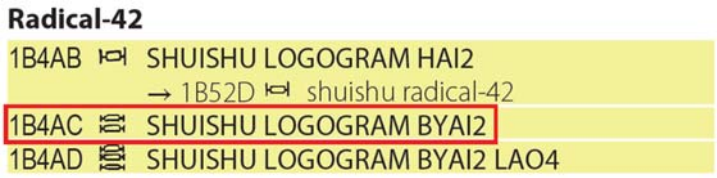
『水書』[9]

numerals in the past researches

Comparison of numerals between proposed Shuishu and past researches.

2. Pronunciations in character names

Comparing with other researchers documents on Shuishu, the stability or generality of phonetic values in the character names of PDAM2.2 Shuishu is slightly questionable. Here I take U+1B3B9 as an example.

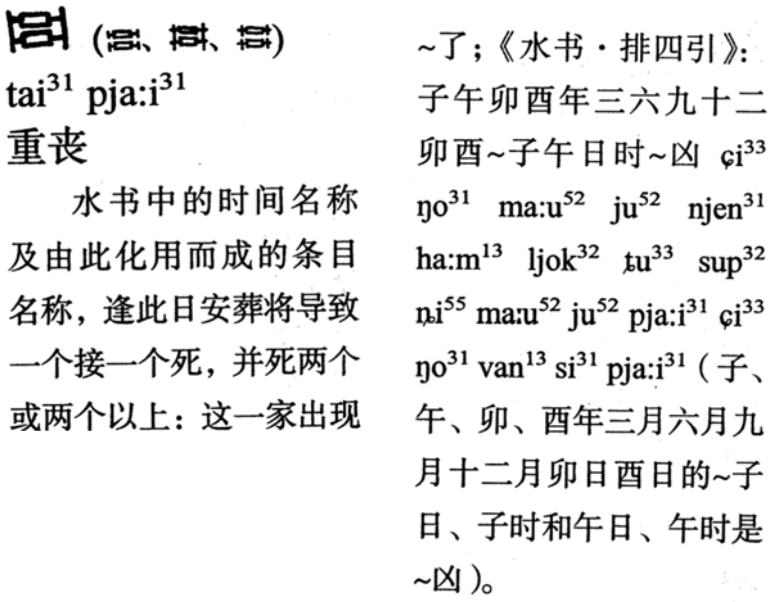


Single, double and triple coffine character in PDAM2.2 [6]

The radical-42 (U+1B52D) is recognized as a coffin, and some big unfortunate is described by repeating it.

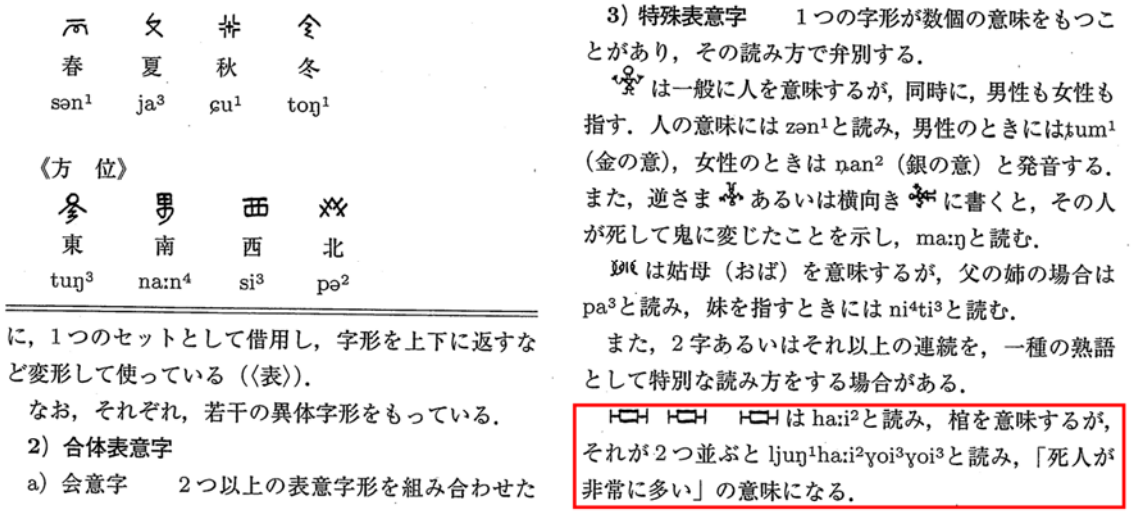
33			27	BYAI2	pja:i^{2}	重丧
34			28	BYAI2LAO4	pja:i^{2} la:u^{4}	大重丧

Semantics of double and triple coffine characters N4758 [3]



Double coffine character in 水書常用字典 [11]

Same glyph with same semantics (see Hanzi annotation) is found in 水書常用字典 [11], but their phonetic values are different. Also, the explanation on Shuishu in “言語学大辞典” [12], by 西田龍雄 (Nishida Tatsuo), shows different phonetics (note: .Nishida explains double coffin is not single character but composite glyph of 2 same characters).



春 夏 秋 冬
sən¹ ja³ ʂu¹ toŋ¹

《方 位》

東 南 西 北
tuŋ³ na:n⁴ si³ pə²

に，1つのセットとして借用し，字形を上下に返すなど変形して使っている（〈表〉）．
なお，それぞれ，若干の異体字形をもっている．

2）合体表意字

a）会意字　2つ以上の表意字形を組み合わせた

3）特殊表意字　1つの字形が数個の意味をもつことがあり，その読み方で弁別する．

は一般に人を意味するが，同時に，男性も女性も指す．人の意味には zən¹と読み，男性のときには ɲum¹（金の意），女性のときは ɳan²（銀の意）と発音する．また，逆さま あるいは横向き に書くと，その人が死して鬼に変じたことを示し，ma:ŋと読む．

は姑母（おば）を意味するが，父の姉の場合は pa³と読み，妹を指すときには ni⁴ti³と読む．

また，2字あるいはそれ以上の連続を，一種の熟語として特別な読み方をする場合がある．

は ha:i²と読み，棺を意味するが，それが2つ並ぶと ljuŋ¹ha:i²ɣoi³ɣoi³と読み，「死人が非常に多い」の意味になる．

Double coffine character in 言語学大辞典 別冊 世界文字字典[11] p.524 (2001)

N4638 [1] wrote "*Its vocabulary is restricted to divination. It contains the knowledge of Shui people on calendar, stars and divining. The number of the words in Shuishu is as small as approximately three hundred*", so it would be incorrect to assume as if a Shuishu character and a spoken word have 1:1 mapping. Even if a single semantics could be identified at an abstract level, it does not mean that we can identify the best phonetic value, as we have many different words for similar idea, like double coffine character.

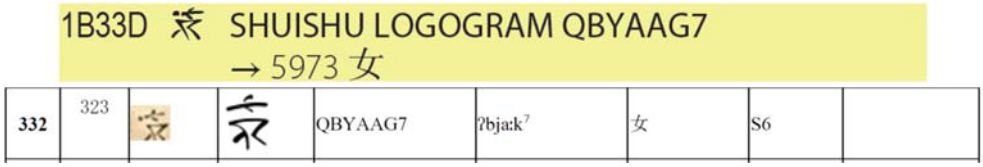
Definition of the authorized pronunciation of Shuishu character would be long term work, it should not be done just for the inclusion of Shuishu character into ISO/IEC 10646.

2.1. Concerns on ASCII notation of Shui language pronunciation

N4758 [3] tells as if authorized romanization method is used in the literacy textbook (see p.3 of N4758), but the quoted 《水族文化进校园读本》 uses a notation without numerals, which is different from PDAM2.2 character names.

《水族文化进校园读本》 quoted in N4758 [3]

By the quotation of conflicting document as if it were supportive of the proposed romanization, there is a concern that few experts had checked the background of the submission during 2 years. In fact, in PDAM2.2 character names, there are initials which are not listed in N4638, N4696 and N4758; U+1B33D has an initial "ʔbj" since which was never listed in the past submission; N4638, N4696 and N4758. Is it a consonant conjunct?



U+1B33D QBYAAG7 and its IPA notation ʔbjaːk³⁵, (upper [6], lower [3])

The user community of the romanization for Shui language would be far larger than that of Shui script, so it is recommended to keep from making snap decision about the romanization, just for Shuishu character encoding.

3. Summary

In summary, this document described 2 points to be resolved for the standardization;

A) the purpose of the standardization is unclear.
   - it is unclear why the strong normalization about the glyph shapes is needed, so either it is unclear how the glyphic variants should be needed.
   - some representative glyphs in the proposal are different from the past researches, it is unclear how the past research documents could be digitized by the proposed character set.
   - in the quasi-statistic method to choose the representative glyph, there might be some undocumented rules to refuse the frequently used glyphs.

B) pronunciations of the characters are really needed in the character names?
   in earlier submissions, the justification of the pronunciations could be rationalized by the ordering of the characters. but PDAM2.2, the ordering is no longer phonetic. still is it needed? what kind of the difficulties there are, if the phonetic values are separated to mutable database in ISO/IEC 10646?
   - some characters are questionable if there is stable reading. in pictographic script, it is popular situation in the pictographic script that single symbol could be red differently by reflecting surrounding context, and there is no stabilized single phonetic value.
   - there are existing documents using different romanization rules, even in the education (if N4758 [3] describes the situation precisely).
   - the romanization rules described in the past submissions ([1]-[3]) are underspecified.

The proposals to resolve these issues are following:

A) please revise the submission of the proposal, with clear purpose. it is strongly suggested to make a self-contained document instead of the short errata of the previous submissions.
   - please provide the raw statistic data, and please write down all rules how the representative glyphs are chosen.
   - please clarify how the variants should be dealt, and consider the possibility of variation selector.

B) please remove the phonetic values from the character name.
   - if therse information should be included in the standard, it should be separated to different database. it makes easier to make a correction in future, and collaboration with the users of different romanization methods.

References

[1] WG2 N4638, “Proposal for encoding Shuishu in the SMP of the UCS”, China NB (2014-08-20)

[2] WG2 N4696, “Preliminary Proposal for encoding Shuishu in the SMP of the UCS”, China NB (2015-10-19)

[3] WG2 N4758, “Updated Proposal for encoding Shuishu in the SMP of the UCS”, China NB (2016-09-03)

[4] WG2 N4839, “Towards the ordering of the Shuishu script in the UCS”, Michael Everson (2017-07-26)

[5] WG2 N4894, “Results of the ad-hoc meeting on Shuishu in Hohhot, 2017-09-21”, (2017-09-22)

[6] WG2 N4922, “Additional repertoire for ISO/IEC 10646:2017 (5th ed.) Amendment 2”, (2017-11-22)

[7] WG2 N4938, “Draft Dispositions of comments on PDAM2 to ISO/IEC 10646 5th edition”, (2018-03-13)

[8] 張均如: 『水語簡志』, 中国少数民族語言簡志叢書, 民族出版社, (1980-11)

[9] 王品魁:『水書』(正七卷壬辰卷), 貴州民族出版社, (1994-12), ISBN 7541204358

[10] 《中國水書》編審指導委員會: 『中國水書』, 四川出版集團 (2006.12-)

[11] 韋世方: 『水書常用字典』, 貴州民族出版社, (2007-09), ISBN 9787541214783

[12] 西田龍夫: 『言語学大辞典 別冊 世界文字字典』, 水書, p.523-526 (2001), ISBN 4385151776)

(end of document)