1. Background

- The four characters whose sequence numbers are 00498, 00500, 00501, and 00502 in Table 7: Glyph and Font (IRGN1921UpdatedComments_Table7.pdf) are shown below.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>00498</td>
<td>00518</td>
</tr>
<tr>
<td>JMJ-058412</td>
<td>JMJ-059347</td>
</tr>
</tbody>
</table>

- ROK suggests not to encode these four characters.
2. ROK’s analysis of the four chars.

- ROK considers that those four characters are derived/modified as follows:

2.1 original char -> cursive form (行書) or handwriting
   -> print (block) form 行書的 楷書化

<table>
<thead>
<tr>
<th>original char</th>
<th>cursive form (行書) or handwriting</th>
<th>print (block) form (楷書化)</th>
</tr>
</thead>
<tbody>
<tr>
<td>與</td>
<td>⇒ 蜣 JMJ-058412</td>
<td>⇒ 蜣 JMJ-058413</td>
</tr>
<tr>
<td>與</td>
<td>⇒ 螟 JMJ-059343</td>
<td></td>
</tr>
</tbody>
</table>

- Problems: In general, several cursive forms could be derived from the ONE original char; furthermore, several print (block) forms could be derived from EACH cursive form. Therefore, in general, tens of print (block) forms could be derived from ONE original char.

2.2 Simplification (簡化): original form -> simplified form

<table>
<thead>
<tr>
<th>original char</th>
<th>simplified form</th>
</tr>
</thead>
<tbody>
<tr>
<td>與</td>
<td>與 JMJ-059347</td>
</tr>
</tbody>
</table>

- In case of Chinese Simplified char, there is a "fixed and stable" correspondence between Traditional and Simplified chars.
- Problems: However, in general, several (or even tens of) simplified forms could be derived from the ONE original char depending on which part to simplify and how much to simplify.
3. Relevant sample data from Taiwan
- 中華民國教育部 異體字字典 : http://dict.variants.moe.edu.tw/
  The search result for ‘與’ in the dictionary is shown below:

4. ROK’s questions

4.1 Does IRG really want to encode all the variants in UCS?
- Which of them can be encoded in UCS and which cannot?
- Can IRG set up rules as to which of them can be encoded in UCS and which cannot so that IRG and experts in general can accept such rules?

4.2 In principle, it is desirable to minimize the number of variant chars encoded in UCS.
- Specifically, it seems that the four JMJ chars can be unified with 與 (U8207).

4.3 This discussion is NOT just about four chars.
- If IRG should allow these four JMJ chars to be encoded in UCS, then hundreds of or thousands of (or even more than 10,000) variants chars may have to be encoded in the future.
  - Is it desirable?
  - Does IRG really want to see such a chaotic situation happen?

4.4 Furthermore, if we encode lots of variants in UCS, then,
- users will have difficulty in choosing the correct(?) one among several similarly looking variants; and
- many existing search systems won't work well.
  - It is contrary to the VERY basic idea of CJK Unification.
  - It implies that information interchange becomes harder and inefficient.

4.5 ROK will probably submit further feedbacks with more relevant data in the future.

***