Update on Auto-Coding – January 2015

When we (the MRC on behalf of HRAF) first engaged with ÜberResearch the intention was to try and assist with coding HRCS Health Categories, hoping that suggested codes would reduce the burden on the members on completing both their first and second pass coding for the 2014 report. However, as the planning progressed it transpired that most members had already completed first-pass coding, and many had even undertaken second-pass coding.

Therefore, in the interests of simplicity for all concerned we have decided to carry on this year's data collection solely via Excel sheets. This keeps things simple and familiar, which we hope will facilitate a successful and timely completion of the dataset collection process and QC coding as discussed in recent HRAF meetings.

Our thanks ÜberResearch for their help in the planning process, and for all their work in accommodating our changing requirements. We will continue to work with ÜberResearch to prepare some upcoming coding tasks and on the analysis part of the exercise. As a result of this continued collaboration, we still intend to provide a final dataset for use by ÜberResearch to further develop the principles and applications for automated coding. The only difference to this upload is that it will be provided as a finalised combined dataset, rather than a pool of un-coded awards for coders to work on. Please note that the Data Usage and Publicity guidance relating to data uploaded to Dimensions remains as previously stated.

By providing this final dataset, we allow ÜberResearch to develop auto-coding in these areas in particular:-

1) Establishing an underlying manual coding 'error rate'. It's important to understand the normal variation of manual coding in order to create a baseline for evaluating non-manual systems.
2) Providing ÜberResearch with third-pass codes (for internal use only) to improve the semantic coding (see point 5 below).
3) Create a short report on the different variation rates between manual and semantic coding.
4) Use semantic codes (with a then known tolerance level) against the wider data-set in Dimensions, so that we can add a deeper analysis within the report about activity in non-HRAF members' data. For example, other RCUK councils, or what is happening in the National Institutes of Health in the USA. Etc. This wider analysis helps to contextualise UK activity, which wouldn't otherwise be possible.
5) "What-if" work. The HRAF have discussed, in the past, the idea of making changes to the HRCS system. The Dimensions system provides us with a platform whereby we can safely assess the impact of any changes before deciding whether to make them. The amended semantic codes would also thread through the back-history.

So our apologies for this late, breaking change, but the practicalities of these changes are easy to make: Excel will be used for recording first, second and (if required) third pass coding. We hope that next time, with improved semantic categories, and a bit more time to prepare, ÜberResearch will be able to support first pass coding for all HRAF members for the HRCS exercise.

Finally, under the old arrangements participating organizations were to get temporary free access to Dimensions. However, now that there is no actual need for users to get into Dimensions, access will now only be set up if you request it. Please feel free to contact ÜberResearch via Giles Radford at giles@uberresearch.com.