HRCS Report 2014: WIP Dataset Submission Guidance

Introduction

As discussed at the last HRAF Meeting (17th Sept), and in subsequent HRCS-related meetings we're keen to have a work in progress (WIP) dataset available from all participating organisations. There are three purposes to this:

1) Begin the second pass Quality Control (QC) coding.
2) Use this data to begin developing the tools required for the analysis.
3) Start the analysis while final data is being collected.

Given there was a lot of discussion at the HRAF meeting and HRCS Training sessions about the format and content of this, can I please ask for this guidance to be used when compiling / formatting the data for submission. This document is intended as a companion to the HRCS Report 2014 FAQ document.

As always, if there are any problems or queries (from HRAF, coders or support staff) please feel free to contact me and I will be happy to assist.

Many thanks,

Jim Carter

Inclusion/Exclusion Criteria (the basics)

The basic requirements for inclusion remain the same as previous reports:

- Research is funded by a participating organisation
- Research must be taking place within the UK
- Research is of health or biomedical relevance

The awards included in this analysis must be active (i.e. incurring spend) in the 2014 calendar year (01/01/2014 to 31/12/2014).

There are a number of elements of funding that are essential to support research activity. These include **direct costs** such as peer reviewed research awards and **indirect costs** such as "administration" and "building maintenance". This analysis focuses on the directly funded peer reviewed UK research of the participating funders.

However, as with the report in 2009/10, we will be carrying out a separate assessment on the level of investment in **indirect** costs. Therefore it is important that ALL awards, whether **direct** or **indirect**, are part of the data submission you provide us. Further details on how to achieve this are shown in the "Indirect Awards" section below.

Calculating Award Value for 2014

We ask all funders to provide **one of two** values for their awards:

- **TotalAward** – the total value of the award for its complete duration.
- **AnnualAward** – the per annum value of the award.

  o Please note this is an either/or option. <u>Do not</u> supply an award value in both Total and Annual formats.
  o We then use the values provided to determine the allocated value for the period the award was active in 2014.
  o Please note we are not asking for actual, budgeted or estimated spend profiles. This will mean the allocated values will not reconcile to other financial / budgeting reports from participating organisations.
  o Currency is assumed to be pounds sterling.

To calculate the award value for 2014, we use an annualised commitment calculation; this means dividing the total value of the award by the number of days/months it is/was active, then multiply by the number of days/months it is/was active in 2014. This gives a predicted amount of spend.

Indirect Awards – Infrastructure & 'non-coded' awards

As with the previous report, our aim is to not use HRCS coding directly for indirect funding such as infrastructure maintenance or miscellaneous 'non-coded' awards, however we do aim to better define what "infrastructure" is. This will be one of the key components of the planned review workshop.

Therefore if indirect/infrastructure awards are part of your dataset you provide us, we would like to be able to classify them more easily. We have therefore included some additional data fields to identify those awards that fall into the areas of <u>Infrastructure</u> or <u>Personal</u> health research related funding, or cannot be coded (<u>Incomplete</u> or <u>Non-health Research</u>). **This is the only column which <u>may</u> need to be filled in manually.** Everything else should be relatively simply transferred from a funder's own records to this excel format. Further information on this can be found in the **IndirectAward** data field notes below.

Otherwise, the inclusion/exclusion criteria remain the same as the previous report. Explanatory text is found in the published report, sections 2 (Scope) and 3 (Methods), pages 18-28. The most recent report (2009/10) can be found on the HRCS website: http://www.hrcsonline.net/pages/reports

If you are still unsure about how to categorise an award, please contact the HRCS Project Management Team.

Data Collection Notes

Each funder will have a different method of award portfolio management and therefore different methods of data export. Therefore we will be asking each funder to provide a short description of their coding approach and participation in the analysis.

This format matches the 2009/10 report's Appendix 1, and <u>we strongly recommend funders familiarise themselves with their previous method of submission as this may help with funder-specific queries.</u>

Data Export Format

- Format is the standard "one line per grant" data entry excel spreadsheet, as found on the HRCS website http://www.hrcsonline.net/pages/analysis:

- Please do not apply 'wrap text' to any columns in the spreadsheet. This may be helpful to you if you are looking at a small/subset of awards, but in a spreadsheet of hundreds, it does not help the analysis.

- Please note that with the additional requirements we have added for this year's report, the spreadsheets are not the same as those used in the 09/10 report. The details of the each data field requirements we have made are outlined below.

If your WIP Dataset is not already coded

We have two primary aims for requesting a WIP dataset at this time:
- To assess the remaining workload by number of awards and proportion first pass coded
- To begin the assessment/assignment of ready-coded awards for second pass QC coding.

Therefore, whether you're portfolio has been already coded or not, please submit anyway.

However, if you already have plans in progress for getting your portfolio first pass coded, please inform us of the procedures and time frame you will be using for this. This information needs to be included in the email you send with your data submission.

Example #1:
"We have contacted an external coder <NAME> who will be providing our first pass coding. They are expected to complete this work and return the data set to us by <DATE>."

Example #2:
"Our internal coding team are working on completing our portfolio's first pass coding, and will have a fully coded dataset by <DATE>".

Data Export Column Headings and Descriptions

A brief summary of the data entry spreadsheet columns are available on the HRCS Report 2014 FAQ document. Here, we've provided further details to help you complete the data export as easily as possible.

In particular, use of the FILLDOWN excel function will be of considerable benefit for all users of the spreadsheet, and we advise all those submitting data to familiarise themselves with this function if they have not used it before. In all likelihood **FunderOrganisation, FunderAcronym, PICountry, CoderName** and **QCCodingApplied** will be the same for all awards, and can be safely filled in this manner.

Finally, please note we are not expecting those submitting the data to spend endless hours filling in all empty fields, especially if the field is not a mandatory requirement. If a required field (e.g. **PIMiddleName**) has data missing on your database, do not worry. We are asking for as complete a submission as possible, but do not intend to create hours of work for those doing so.

FunderOrganisation, FunderAcronym and OrganisationReference (Columns A, B & C)

With the exception of AMRC, all funders submitting data will be from a single organisation. Therefore your funder name and acronym should be the same for all award lines.

The **OrganisationReference** is the unique identifier you use to track the award in question. This is more commonly known as the 'grant code', 'grant ID' or 'award#' depending on your organisation. If awards only have a numerical ID, then please add your **OrganisationAcronym** as a prefix, e.g. MRC1234. This prevents confusion if two funders share an **OrganisationReference** with the same numerical ID.

PI Details; Title, Name, Address, etc. (Columns D to O)

These details are required for both the HRCS project and UberResearch for similar reasons; to allow de-duplication and disambiguation screening. Data publicity is discussed at length in both the FAQ and Publicity Satement, but in short the HRCS Report will not include PI name in the publically available dataset. **PIInstitution**, **PICity** and/or **PIPostCode** may be used to assess distribution of funds across UK regions, and will be in the public dataset.

PIInstitution should also be the overall host organisation. An embedded unit or centre within an Institution would be used as **PIAddressLine1**. For example, a PI at the MRC Toxicology Unit at the University of Sheffield would have "University of Sheffield" as **PIInstitution** and "MRC Toxicology Unit" as **PIAddressLine1**.

Funding Mechanism & FundingStream (Columns P & Q)

Establishing the mechanism and stream of funding will be helpful in further HRCS report sub-analysis. However this is not necessarily recorded by all funders, and even if there is such a field available, the definitions used may not be the same from one funder to another. As such, neither is a mandatory data field. If they are available in your database, please include them. If not, move on.

StartDate, EndDate, Duration (Columns R, S & T)

Both start and end dates need to be in DD/MM/YYYY format. Duration needs to be months expressed as an integer. By using this format, it will allow some automated excel formulae to confirm this has been calculated correctly. You'd be surprised how often end dates can erroneously appear before start dates!

This also allows for correct calculation of the 2014 award amount, based on the dates/duration and **TotalAward/AnnualAward** value. Other date formats could be supplied by arrangement, but this is our preference.

TotalAward / AnnualAward (Columns U & V)

If a total award amount is available please populate the **TotalAward** field. If funding is allocated on an annual basis only, please supply the amount allocated for the 2014 calendar year as **AnnualAward**. Currency is assumed to be pounds sterling.

Please note this is an either/or option. Do not supply an award value in **both** Total and Annual formats.

AwardTitle, AwardAbstract and Keywords (Columns W, X & Y)

Award Title and Abstract need to be made public to allow any subsequent sub-analysis of the dataset, e.g. by keyword search. However, if these fields contain sensitive information, such as personal information, details of *in vivo* experiments, patent details or copyrighted materials we will ensure the data is anonymised.

Please note that **AwardAbstract** refers to scientific abstracts only. Lay abstracts will not be suitable for coding purposes unless there is no scientific abstract available.

If your organisation's database includes a column for Keywords, please supply it here. This is non-mandatory as many funders may not have this information available or able to export in this format. However anything that is provided could aid further sub-analysis of the dataset.

IndirectAward (Column Z)

As explained earlier, the HRCS Report and Analysis has a primary focus on <u>direct</u> research funding, principly grants with a specific research aim. However a significant proportion of Health Research funding goes towards <u>indirect</u> costs, such as the buildings, administration and personal awards.

To segregate between direct and indirect awards, we are asking funders to fill in this field, **IndirectAward**, for any awards that fit one of the four criteria listed below. Please use these specific terms **in bold** (and only these terms) for the following options:

- *<Leave Blank>*
    - o For <u>direct</u> awards that have been coded as normal.
- **Infrastructure**
    - o The award falls into the umbrella term "Infrastructure". To broadly distinguish from "Personal", "Infrastructure" awards can be described as those funding a site or group, whether it is in administrative, construction or maintenance. Where possible, you must also use the "Coder Comment" (Column E) to add further details.
    - o Examples of "Infrastructure" awards include:
        - ▪ Buildings & equipment (construction, maintenance)
        - ▪ 'Core' funding (e.g. for MRC units) & other administrative costs
        - ▪ co-ordinated ventures like CRNs, Methodology Hubs etc..
        - ▪ Awards to support hosting/organising conferences
- **Personal**
    - o The award <u>IS</u> health research related, is <u>NOT</u> infrastructure but <u>may</u> not 'codeable'.
    - o To broadly distinguish from Infrastructure, "Personal" awards are those that are given to an individual for a specific purpose such as training, salary support or conference attendance.
    - o Again, any details for this can be included in the "Coder Comment" section.
    - o Examples of "miscellaneous" awards include:
        - ▪ Travel grants and other awards for conference attendance
        - ▪ Bursaries and other awards to support researchers directly
        - ▪ Professional membership fees
        - ▪ Funding for clinician training, lab visits and secondments
- **Incomplete/Missing**
    - o There is not enough information to be able to judge what the award is for. E.g. if title/abstract is missing and/or unclear. Again, please include details of why the grant is classified as "Unknown" in the "Coder Comments" section.
- **Non-health research**
    - o The award is <u>definitively not</u> health research related.
    - o In most cases, we would not expect to see these awards at all, but if they are included in your data submission we need to be aware how to handle them.

CoderComments (Column AA)

Any other comments or caveats relating to a particular grant should be added to the "Coder Comment" column.

<u>This includes any explanatory details relating to the **IndirectAward** options (Infrastructure, Personal etc.) above.</u>

CoderName (Column AB)

If it is possible for you to trace coding to specific person, please put their name/initials here.

This allows us to ensure we send awards for second pass coding to another independent coder for blind coding. If your portfolio is only coded by one specific person, use the fill down options to ensure all awards have a coder name associated with them.

If your organisation has multiple coders, and you cannot trace which coder worked on which award, please leave this column blank.

QCCodingApplied (Column AC)

We are aware that a few funders already have their own internal QC procedures. If this applies to you, please indicate this by putting the word "Yes" in this column.

In most cases the record of specifically which awards have been checked will not be recorded. Therefore we ask all those submitting data that has been QC'd to let us know the process involved and the proportion to which it has been applied. As you know, we are aiming for a 40% overall QC rate for the combined data.

If methods of second pass QC coding varies between different divisions of your organisation, please keep those datasets separate when submitting (if possible). This can be by including them as separate worksheets within the same excel file, or as individual excel files, whichever is preferable to you. All we require is a note from you when you submit them to indicate the methods used for each sheet/file. We can then apply appropriate second pass coding based on this information for the report submission.

AssignedUberUser (Column AD)

This is used by the project management teams only. By submitting the spreadsheet to UberResearch with an UberResearch username (coder's email address) attached to an award, the award will be automatically assigned to that user for QC coding, thus simplifies the distribution of the communal coding pool.

Providing Further Updates

There will be several reasons for updating the dataset;

- Providing data on further awards active in 2014 (e.g. awards that have started since the last update)
- Providing first or second-pass HRCS codes for awards.

The WIP dataset will be an incomplete picture of all awards active in 2014, so clearly this will need to be updated by early 2015 with the remainder of your portfolio data. Please do bear this in mind and consider carefully how you will identify the remainder of these awards ASAP.

Our preference is to "top up" the dataset with awards not previously notified to us, but if you can take particular care to ensure that all awards have a **OrganisationReference** unique identifier then we will routinely scan for duplications.

If the update is to add HRCS codes (either first or second-pass codes) then please also make this clear on the spreadsheet cover.

The planned deadline for the updated, and final, dataset is Friday the 13th of March 2015. At this point we should have all awards collated, coded and our target of 40% second pass coded.

If you have any other problems or queries

We welcome any specific problems or queries you have regarding the data submission and coding process. The more feedback we receive, the better the guidance we can provide, making the report more consistent in how the coding is applied and analysed. Any specific examples will be added to **Annex 1** (see below).

For all queries, comments or for general help regarding the HRCS Project, please contact Dr. Jim Carter, the Project Manager for this report on 0179 341 6240 or james.carter@headoffice.mrc.ac.uk.

Thank you for your time.

Jim Carter

Annex 1 – HRCS Queries Log and Specific Award Guidance

How do I handle award extensions?

Sometimes funders grant additional funding to an established award, to extend the duration or provide specific funding for additional resources. In these cases the title/abstract for both original (parent) and extension (child) awards will be the same/very similar. The awards may also be linked in some fashion within a funder's database.

The optimal method for reporting this is to combine the additional funding to the original award, and apply HRCS coding only once to the combined award. If the further funding changes the lifetime of the award, make sure this is incorporated into your **StartDate**, **EndDate**, **Duration** and **TotalAward**/**AnnuallAward** values.

If this combined award is too difficult to edit within your portfolio / spreadsheet submission, our main requirement is that both parent and child awards <u>MUST BE</u> coded identically, and note made in the **CoderComment** column to indicate that the two awards are linked.

There are two exceptions to these rules:
- If the additional funding falls into the category of **IndirectAward** (e.g. if the extra funding is for administrative support, not research) – see the guidance on this for more details.
- If the additional funding is for a new area of research that would have to be coded differently.

In both these situations the title/abstract should make this clear, and will in all likelihood be different to the original. In these cases, the extension is handled as its own separate award and therefore you should follow your normal procedures for coding. It would still be useful to indicate the link between the two awards in the **CoderComment** column, but this is not mandatory.