"The classification of research portfolios"

Member Organisation Forum on Publicly Funded Research
Working Group on "Comparative Research Portfolios"

Contents

Executive Summary

There is growing interest in the national and international alignment and co-ordination of funding strategies, as Governments seek to explain and maximise the return on investment from public funding in research and development. Funding agencies also wish to be able to: identify gaps and opportunities, avoid duplication, encourage collaboration and strengthen the case for research.

A stable, consistent and appropriately detailed classification system for analysis of research portfolio information, can support these objectives, and is a solid foundation for the evaluation of progress, productivity and quality. A working group of the ESF Members Organisation Forum on "Evaluation of Publicly Funded Research" met to discuss approaches for categorising research portfolios, with the aim of exchanging expertise and experience in this area.

This report summarises the meeting discussions, which largely focussed on the application of the Health Research Classification System (HRCS). The HRCS was of interest as it is an example of system used internationally across a number of organisations for the analysis of medical research portfolios. The working group acknowledged that other approaches are used in other disciplines, and many funding agencies have developed in house systems. Barriers to developing common approaches for categorising research portfolios were examined.

Key points detailed in the report include the need to ensure consistent and high quality classification, how the wider use of standard classification systems might be encouraged (for example across the diverse research organisations in Europe), methodological suggestions regarding improvements to the efficiency/range of analysis that can be performed, and examples of how portfolio analysis can support strategy development.

The report highlights that this is an important area of work and that progress at the policy and practical level across Europe will help support the ESF/EUROHORCS vision for a globally competitive European Research Area.

1. Background

1.1. ESF MO Forum Working Group on classification of research portfolios

A number of organisations within the ESF Member Organisation (MO) Forum on "Evaluation of Publicly Funded Research"[1] established a Working Group[2] to exchange experience and expertise in two important areas of research management; the analysis of information about research portfolios, and the collection and analysis of output data. After preliminary discussion of this area at the ESF MO Forum Workshop in Slovakia (November 2010), the working group agreed to meet with a wider set of representatives from research organisations in London (December 2010).

The Forum is composed of 33 ESF Member Organisations (MOs) and 7 observers, and was established following the need to exchange information on evaluation practices and practical approaches within MOs. This Forum is the sole dedicated platform for European research funding and performing organisations that provides a continuous forum to exchange information and work together on common projects. So far the Forum held two workshops (Stockholm spring 2010 and Slovakia November 2010) and it will finalise its activities during 2012.

Members of the working group were interested in the analysis of research portfolio information in order to:

- Monitor progress against organisation strategy (e.g. consistently reporting investment over time by scientific area)
- Organise the peer review process (assign reviewers, and allocate applications to research boards)
- Communicate balance of funding (e.g. demonstrate which areas and types of research are receiving investment)

There is growing interest in the national and international alignment and co-ordination of funding strategies, as Governments seek to explain and maximise the return on investment from public funding in research and development. Funding agencies wish to be able to: identify gaps and opportunities, avoid duplication, encourage collaboration and strengthen the case for research. It was agreed that a stable, consistent and appropriately detailed classification system can support these objectives. Members of the working group wanted to emphasise the advice to:

- keep approaches as simple as possible
- maintain existing processes
- start at a high (aggregated) level of analysis
- ensure consistency (between funding agencies and over time)

It was identified that detailed discussion was needed on the following issues:

- Current approaches for the classification of research portfolios (In particular whether processes for categorisation were researcher-led, internal or automated)
- How quality was assured (is this achieved by sampling, training coders, automating the process or a combination of these?)
- What are the approaches organisations use for analysis and publication of results?

---

[1] ESF MO Forum on Evaluation of Publicly Funded Research http://www.esf.org/activities/mo-fora/evaluation-of-publicly-funded-research.html

[2] ESF MO forum on evaluation Working Group current membership: Brendan Curran (Ireland, Health Research Board) , Katharina Fuß (Germany, DFG) Iveta Hermanovská (Slovakia Academy of Sciences), Katrin Milzow (Switzerland, Schweizerischer Nationalfonds), Jenny Nordquist (Sweden, Swedish Research Council), Ingrid Roxrud (Norway, The Research Council of Norway), Ian Viney (UK, Medical Research Council), Rafael de Andrés Medina (Spain, Instituto de Salud Carlos III)

1.2. Approaches to the classification of research portfolios

A classification sorts data about research funding into discrete categories. Many classification systems follow a structure which branches from high level categories to more specific sub-categories, allowing increasingly more detailed analysis. This is the fundamental first step to understanding resource flows into research and development. Maintaining a consistent approach to categorisation allows changes in research portfolios to be tracked over time.

Within the ESF MO Forum working group a variety of classification approaches are used. Most organisations utilise bespoke internal systems and some had additionally adopted systems that were used by a number of other organisations (e.g. the Health Research Classification System, Frascati[3], and MeSH[4]). The variety of systems in part reflected the fact that there were several purposes for such classifications. In some cases internal systems are highly integrated with administrative processes across the organisation, for example driving the allocation of applications to research boards (e.g. the DFG system explained in the presentation at **section 2.5.**). In addition systems support the organisation's ability to answer questions about the amount invested in specific scientific areas, or on different types of research.

Across these approaches there were a number of properties agreed to be important, although not all organisations wished to necessarily have all these in a single classification system. It was acknowledged that organisations, particularly those with a large and/or diverse portfolio, may need to employ several systems to address all their operational, policy and strategic requirements. Research organisations selected classification systems that support analysis at an appropriately granular level, and separated categorisation of the stage/type of research from subdivisions of disciplines or subject areas.

The systems in use across research organisations included the Health Research Classification System (HRCS) and the OECD Frascati manual. Together the working group members had greatest experience in applying the HRCS to analyse their research portfolios, and so discussion was largely focussed on how this system could be applied and improved. It was acknowledged that this meant the working group focussed their discussions on the application of classification approaches in a single disciplinary area, and largely concentrated on a single categorisation approach, however it was expected that discussion would illustrate issues of wider impact. Members emphasised that the ESF MO Forum may wish to consider implications across other disciplines, it was also noted that the European Medical Research Council planned to look across a wider set of systems in order to consider a recommended approach for classifying health research in Europe.

1.2.1. The Health Research Classification System (HRCS)

The HRCS is a system for classifying and analysing biomedical and health research funding. Its role is to facilitate research management by answering strategic questions about investment. The HRCS is a two dimensional framework for classifying health research[5]. Codes from both HRCS dimensions are applied when classifying:

- One dimension, the **Health Categories**, is used to classify the type of health or disease being studied. There are 21 categories encompassing all diseases, conditions and areas of health.

- The other dimension, the **Research Activity** Codes, classifies the type of research activity being undertaken (from basic to applied). There are 48 codes divided into eight

---

[3] OECD Frascati Manual:
http://www.oecd.org/document/6/0,3343,en_2649_34451_33828550_1_1_1_1,00.html
[4] MeSH stands for "Medical Subject Headings" and is the National Library of Medicine's controlled vocabulary thesaurus. It consists of sets of terms naming descriptors in a hierarchical structure that permits searching at various levels of specificity. More information can be found at
http://www.nlm.nih.gov/mesh/
[5] Full information about the HRCS can be found at www.hrcsonline.net

groups: 1 Underpinning, 2 Aetiology, 3 Prevention, 4 Detection and Diagnosis, 5 Treatment Development, 6 Treatment Evaluation, 7 Disease Management and 8 Health Services

HRCS codes are assigned to capture the main objective(s) of a particular study - so the system provides a broad overview of the *centre of gravity* of a set of research awards.

Defined percentages are assigned to all HRCS codes - which means that the associated funding is analysed exactly with no double counting. In order to classify research projects using the HRCS a short scientific summary of the proposed work is required, usually suitable text is available within application forms to do this.

The strategic aim of coding using the Health Research Classification System is to capture the main objective of the research taking place during the lifetime of the award and not the background or future potential downstream applications of the research.

1.2.1.1. Use of the HRCS in the UK

The HRCS underpinned two important reports issued by the UK Clinical Research Collaboration (UKCRC) which together provided a comprehensive overview of non-commercial health research funding in the UK in 2004/05.

The *UK Health Research Analysis* report, published in 2006, was the first ever national analysis of UK health research. It provided an overview of all types of health research activity across all areas of health and disease in the UK, funded by the 11 largest government and charity health-related research funders. The HRCS allowed meaningful comparisons to be made across the different funders' research portfolios. The report includes[6]:

- A breakdown of spending on all types of health research (from basic to clinical) across all areas of health and disease
- Details of the distribution of funding within individual areas of health and disease
- The geographical spread of health research investment across the UK.

Subsequently the HRCS was used to analyse the funding activities of 29 medium and smaller sized members of the Association of Medical Research Charities. The UKCRC report *From Donation to Innovation*, published in 2007 included an in-depth analysis of research funded by medium and smaller sized members of the Association of Medical Research Charities (AMRC).

The two reports have been disseminated widely in the UK and had a major impact, providing the basis for high level strategy discussions and informing a number of joint funding initiatives.

The main outcomes from the implementation of the HRCS in the UK include:

- A strengthening of the evidence base for strategic discussions by research funders and importantly *between* funding agencies.
- Supporting a clear view from Government about the priorities for medical research
- Stimulating a number of new joint funding initiatives between funding agencies to address particular gaps, opportunities and areas that need capacity building:
  - National Prevention Research Initiative (more than £30m committed via four phases from 16 funders)
  - Public Health Initiative (£20m to fund 5 centres of excellence from 8 funders)
  - Translational Infections Research Initiative (£16.5m for new grants from 7 funders)
  - UK brain banking strategy (appointment of national director and greater co-ordination of activity)

---

[6] Annex 2 includes examples of the analysis presented in the UK Health Research Analysis report

The main public and charitable funding agencies for health research in the UK plan a new analysis of health research using the HRCS, based on expenditure in 2009/10 and for publication in 2011.

1.2.1.2. What are the main advantages of the HRCS

The main benefits of the HRCS are:

- It covers the full spectrum of biomedical and health research - from basic to applied - across all areas of health and disease
- It directly links funding to research objectives
- It has a 5 year history of being used to classify thousands of research awards of different types from many funding organisations and settings
- It is a common stable system allowing meaningful comparisons between different portfolios
- It is open source and freely available, and there is a growing body of guidance and advice to support its use

Some notable aspects of the HRCS are:

- It was designed for a specific purpose - to provide a broad overview of the strategic centre of gravity of a set of research awards
- The HRCS was designed for analysis of health research, although it may be possible to extend the approach to other areas of research, organisations that fund non-health-related research would need to apply other systems to the remainder of their portfolio
- It is based on the main research question to be answered within the lifetime of the award. Therefore it does not capture all facets of the research or potential downstream outcomes and hence does not replace other keyword systems which enable searching and finding. Many research organisations find that they need to operate more than one classification system to address all the questions they may be asked about their portfolio, no approach to classification will be ideal for all tasks.
- It captures costs directly tied to a research aim and therefore excludes indirect background support for research (where there is no associated research question). This means it is not a financial audit tool

The HRCS is an analytical framework where assigned codes are directly linked to award funding using a standardised coding and apportionment system. This means that amounts of investment associated with specific areas of research can be analysed in a reproducible manner with no double counting of total award funds.

A number of simple tools have been developed to enable analysis using the HRCS, including spreadsheets to facilitate data entry and analysis, as well as template databases. This workshop includes an explanation of these tools (presentation at **section 2.4.)**

1.2.1.3. Is the HRCS used outside of the UK?

Since publication of the two analysis reports, the HRCS has been adopted by further research funders in the UK and internationally to inform research management and to undertake prospective analyses. The information provided by the analyses and internal use of the classification system has enabled organisations to examine their own research portfolios in detail.

In the last five years the HRCS has been used to analyse research portfolios in Ireland, Norway, Sweden, Singapore, Hong Kong and Canada. This workshop session included presentations on the work in Norway (**section 2.1**), and Sweden (**section 2.2**).

The HRCS is based upon a classification approach used by the US National Cancer Institute[7] (the Common Scientific Outline, CSO) to specifically classify its research portfolio. It is of

---

[7] International Cancer Research Portfolio (ICRP): http://www.cancerportfolio.org/cso.jsp

interest to understand how the CSO has expanded successfully to support researchers and policymakers worldwide in the cancer research field, as an indication of what the wider use of the HRCS could deliver for the broader area of all health research. The workshop included a summary of progress from ICRP (presentation at **section 2.3**).

1.2.2. OECD Frascati Manual[8]

Within OECD countries research and development investment is classified according to the Frascati manual which guides surveys in this area. Frascati provides a framework for the routine collection of data on research and development in all areas leading to the tracking of expenditure in this area in all OECD countries. Members also highlighted that the Australian and New Zealand Standard Research Classification (ANZSRC), a system produced jointly by the Australian Bureau of Statistics (ABS) and Statistics New Zealand (Statistics NZ) has been successfully implemented in Australia/New Zealand and is based upon the OECD Frascati manual approach.

The Manual is accessible at:
http://www.oecd.org/document/6/0,3343,en_2649_34451_33828550_1_1_1_1,00.html

---

2.   Workshop Presentations

The second workshop of the ESF MO Forum on Evaluation of Publicly Funded Research took place in Smolenice 29-30 November 2010. During this meeting the Working group on Classification of Research Portfolios discussed and summarised experiences of using classification systems to analyse their research portfolios, the advantages and disadvantages of using these approaches, and how they have been used to develop strategy.

The group held a follow up meeting on the December 2010 in London, attended by 16 representatives from funding organisations. The majority of attendees were from funders of medical research and the focus was on using the HRCS as a common language for portfolio analysis. However, the discussion also covered the mapping of the other classification systems to the HRCS (e.g. DFG disciplinary classification system and the Common Scientific Outline used by the US National Cancer Institute and many other cancer research funders internationally). The main issues covered by the workshop's presentations are outlined below.

2.1.   Implementation of the HRCS in Norway (Ingrid Roxrud, The Research Council of Norway)

There are three main sources of public funding for medical and health-related research in Norway. The Norwegian Ministry of Health and Care Services provides research funding for the hospital sector, directly through the hospital budgets (which typically allocate between 1 and 3% of their budgets to research and development) or via the four regional health authorities. The Ministry of Education and Research funds universities. The Research Council of Norway is the main national competitive arena for research funding, it receives funding primarily from the ministries, and grants support to research-performing institutions such as universities, hospitals, research institutes or to the private sector.

In total public funding for health research in Norway in 2008 was approximately 4.8 billion kroner (£520m, €600m).

The Ministry of Health and Care Services in Norway has been instrumental in encouraging a strategic look at what health research is supported in Norway and the use of the HRCS to classify projects has been central to this.

The regional health authorities have classified all research that they supported from 2009, and now the Research Council of Norway plans to begin to use the HRCS to examine its portfolio. The coding system for medical research currently used by the Research Council is fairly similar to HRCS with a research activity type code and then 37 disease categories.

The Research Council will begin a pilot of HRCS coding shortly. This will involve taking a selection of research projects receiving funding after this year's application process. The coding will be done by several groups of people in parallel – project leaders, scientific member of the application committee, programme managers and/or external coders. The process will then be evaluated, and will include looking at the validity and consistency of coding, with the aim of suggesting the most reliable and efficient way to code applications.

Discussion points:

2.1.1. The validation/consistency of coding is very important, in order for information to be shared between organisations and conclusions drawn about relative investments.

2.1.2. Of course considerations of the resources required to classify projects are also important. Currently organisations are approaching classification in quite different ways, and so the work in Norway will be very informative with respect to recommending whether coding can be carried out by applicants, or whether

classification should be checked by research administrators, and the level of checking that is required to achieve consistent coding.

2.1.3. Attendees discussed the fact that for high level codes there was less likely to be large variability between coders, but consistency at the subcategory level was more difficult to achieve.

2.1.4. Currently there is no best practice with respect to processes used to assure adequate quality of classification, each organisation had its own approach to cross check coding, some taking a sample of awards to re-code and some routinely using at least two opinions to code all awards.

2.1.5. It was particularly important to monitor the use of the more general categories (e.g. "generic health relevance") as high use of these may indicate less accurate coding, and lower quality portfolio information overall.

2.1.6. A pragmatic approach for joint analysis is for organisations to be transparent about their approach to quality assurance and quality control. It was suggested that reports including joint portfolio analysis should include a statement from each organisation regarding how they assure awards are classified to an appropriate standard, including the processes for checking coding and training of coders.

2.2. Implementation of the HRCS in Sweden (Jenny Nordquist, Swedish Research Council)

The Swedish Research Council (VR) supports all areas of research with the largest proportion of its funds being allocated to natural and engineering sciences (1.3 billion SEK, £130m, €150m) followed by medicine (1 billion SEK, £96m, €110m).

In 2008, Universities in Sweden received approximately 27 billion SEK (£2.6 billion, €3 billion) for research and research training. 9% of their funding originated from VR, almost half (46%) from central government. FAS (Swedish Council for Working Life and Social Research), Formas (Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning) and VINNOVA (Swedish Agency for Innovation Systems) accounted together for approximately 5%. Swedish scientists at universities and colleges can also apply for funding from Nordic and European institutions, for example, NordForsk (Nordic Research Board), the European Research Council (ERC) and the European Union's Seventh Framework Programme.

VR have initiated two pilots using the HRCS approach. In 2009 VR funded projects from 2007 and 2009 were coded by the chairs of the evaluation panels. In 2010 all proposals have been classified by all panel members. VINNOVA also plans to code its portfolio, and work is underway in Western Sweden to classify government funded projects.

The profile of VR funded research as analysed using the HRCS is similar to the UK MRC with the majority of research supported in the areas of underpinning and aetiology, and the main health category areas being neurological, cancer, infections and metabolic disease.

By reference to WHO data on DALYS (Disability Adjusted Life Years) the portfolio was compared against the relative disease burden in Sweden. As in the UK, disparities were seen between the areas such as infection where the investment in research is relatively high and the burden of disease relatively low.

As all VR applications were coded, not just those that were awarded, VR could examine application success rates by portfolio area. Of interest was that the approval rate for proposals in underpinning research was significantly higher than that for proposals overall, the approval rate for proposals in prevention research were comparable to other areas, whereas the success rates for proposals in disease management and health services research were lower. When success rates by health category were examined areas that were approved at a higher rate included cancer, neurological and stroke research (although there were a low number of

applications in the area of stroke), and at a lower rate included cardiovascular and musculoskeletal.

Next steps for the work at VR is to finish the analysis of the 2010 data, work with the other Swedish funding organisations to get a more complete picture of research in Sweden, and also to evaluate the classification process.

Discussion points:

2.2.1. Attendees congratulated the speaker on VR's implementation of the HRCS approach, which was clearly yielding information helpful to strategy development and decision making.

2.2.2. Attendees continued to discuss approaches for the validation of coding. The example from Sweden raised interesting issues, as none of the coders were trained in the use of the HRCS, but had found it reasonably straightforward to apply the classification system with the standard guidance on use of the HRCS. There might be the opportunity to compare consistency between panel members to see if the guidance could be improved in order to strengthen consistency of coding.

2.2.3. Attendees asked whether the results been fed back to the decision making panels to see if the analysis will have an effect on future funding patterns. The results have been discussed with the VR panel that decides on the apportionment of funding to each of the peer review panels, and so may influence strategically the budget that each area has in future.

2.2.4. It was noted that with low numbers of applications in some areas it was difficult to determine whether the approval rates indicated a lack of quality in these areas, but that with more data a useful picture should emerge.

2.3. International Cancer Research Portfolio (ICRP) (Lynne Davies, Cancer Research UK)

The HRCS is based on the Common Scientific Outline (CSO) system developed by the US National Cancer Institute (NCI)[9] in conjunction with the Department of Defence (DOD) US Army Medical Research and Materiel Command in the late 1990s. The CSO was developed via workshops in which scientists, panel members, applicants and programme staff categorised abstracts, and then evaluated the validity of the coding.

The CSO is now directed and managed by the International Cancer Research (ICR) Partners. The Partners include two national partnerships, the NCRI(UK) and CCRA (Canada), 8 individual funders from the U.S. including NCI/NIH and 2 other European funders, totalling 48 funding organizations worldwide. The Partners meet annually and via teleconference periodically throughout the year to share information and review the implementation of the CSO in their organizations.

Since the formation of the ICR Partners (ICRP) in September 2000, partners have focused on implementing the CSO to meet the needs of all member organizations. The CSO is now being used in many ways to inform internal and joint policies and it provides an internationally regulated framework that ensures comparability, consistency and accuracy of coding.

The ICRP's mission is to add value to cancer research efforts internationally by fostering collaboration and strategic co-ordination between cancer research organizations. As well as joint analysis of cancer research activity using the CSO the ICRP facilitates networking to improve collaboration and co-ordination and the sharing of resources, such as evaluation tools.

---

[9] International Cancer Research Portfolio (ICRP): http://www.cancerportfolio.org/cso.jsp

The CSO is organized around seven broad areas of scientific interest in cancer research, further divided into a total of 38 subcategories (the full categorisation is at http://www.cancerportfolio.org/cso.jsp).

Funding has recently been obtained via an ERA-NET award for translational research (TRANSCAN), for analysis of the European cancer research funding portfolio at the grant level using the ICRP's CSO classification system. 25 cancer and medical research funding organizations are involved in total. The plan is to begin with a short survey and then proceed to detailed grant level analysis. 12 countries are involved in this work package. This will build on the previous work of EUROCAN and ECRM, but aim to create a coded European database of cancer research awards. ICRP will be used instead of duplicating activity in Europe, and work is expected to start in March 2011.

The data provided by ICR Partners are shared on the ICRP website, which was launched in 2003. The web site is provided as a public service by the Partners, and is hosted on the United States Government computer system at the NCI. It is planned to re-launch the website early in 2011 in order to enhance the functionality provided to ICR partners and the public. The public site allows users to search for cancer research awards using defined criteria and is a valuable tool for researchers to identify potential collaborators worldwide   The ICR Partner-only analytical tools on the new web site will allow organizations to conduct their own analyses of the international portfolio, giving Partners an international perspective to help inform strategic planning.   In addition the site will provide online networking tools for Partner organizations via a web site forum.

ICRP has developed a number of policies in order to support the joint analysis of data:

- All Partners have signed data sharing agreements to allow each other access to the full dataset for analysis
- Partners may use the data made available through the restricted view for internal purposes only
- If Partners wish to publish any part of the data/reports on the restricted site, they must gain the approval of the Partnership first.
- Caveats relating to Partners' data are included on the site for reporting.
- New Partners cannot gain access to the international portfolio until they have contributed data and agreed to abide by the policies and procedures.

Partners contribute data as and when it is available, the ICRP database includes information on awards from 2000 and annual analysis is possible from 2005 onwards.

ICRP is keen to collaborate where possible with users of the HRCS system. As the HRCS was developed from the CSO it follows a similar pattern. Attempts to map the systems to each other to allow cross-talk between them have had some partial success; some HRCS codes have a 1:1 relationship, allowing automated translation from one system to the other (e.g. HRCS 7.2 corresponds to CSO 6.6 – end of life care). Other HRCS codes can be combined into a single CSO category (e.g. HRCS 5.4 and 5.5 both map to CSO 5.1 (localized therapies). Overall 32 of the 48 HRCS codes can be automatically mapped this way (67%).

Many HRCS codes have a more complex relationship to the CSO. For example HRCS 2.2 relates either to CSO 2.1 or CSO 2.3. MeSH keyword filters have been examined to see if there are simple ways in which research can be split into different CSO categories. Initial results suggest that this could simplify the work but that some manual intervention would still be required. Similarly, many CSO codes have a simple relationship to the HRCS, but others do not (e.g. CSO 2.3 splits between HRCS 2.1 and 2.2).   Again ways to simplify the mapping process have been examined.

Discussion points:

2.3.1. The ICR partnership demonstrated the benefit of sharing data using a consistently applied and stable categorisation approach.

2.3.2. Attendees noted that with commitment from all funding agencies involved the ICRP database represented a significant central resource which returned considerable added value to all partners. If the HRCS was to be more widely adopted a similar central dataset could be very useful to participating research funding agencies and research performing organisations.

2.3.3. Following the processes and policies agreed by the ICRP, to support joint analysis, could also benefit organisations using the HRCS.

2.3.4. Validation/verification of coding was raised again in response to this presentation, and the importance of capturing any learning about issues in certain areas should be fed back into training material and available guidance.

2.3.5. ICR partner organisations used a variety of approaches to categorise their portfolios. Overall internal programme staff that had opportunity to be trained in applying the categorisation gave more consistent results. Some organisations coded awards, some applications.

2.3.6. Even without developing a central database, the HRCS online website would benefit from having a forum, which organisations using the HRCS could use to exchange information and improve guidance and training material.

2.3.7. Attendees were interested in the use of MeSH coding to assist in translating CSO codes to HRCS codes and vice versa it was agreed that there was value in sharing any learning that emerges from work to map classification approaches.

2.4. Portfolio analysis using the HRCS online tools (Andrew Speakman UKCRC)

The HRCS website contains reports, training information and analysis tools useful to any organisation wishing to use HRCS codes and carry out portfolio analysis. http://hrcsonline.net/

The website was designed to:

- Make all existing information and resources accessible
- Provide guidance for naïve users wanting to learn how to use the system
- Act as a reference source for experienced users
- Provide further contextual help, summaries and linkages

The overarching aim of the site is to promote sustainability of the system.

The analysis tools available to download from the site are designed to reproduce the analysis which took place as part of the two published UK health analysis reports. The analysis is divided into three stages:

1. the process to collect coded portfolio information
2. a database for storing this data
3. an output spreadsheet for graphing the results

1. The data entry spreadsheet is a simplified version of that used in the main UK health analysis, it is compact and easy to edit directly and provides simple functionality for export. The spreadsheet incorporates two basic validation checks:

- It restricts coding choices to valid HRCS codes (In HC_n and RA_n columns)
- It restricts associated percentages to common values (i.e. 33.3%, 50.0%, 66.7%, 100%) but note that it it does not ensure these values add to 100%

The spreadsheet can be easily customised to support more detailed analysis (additional classification columns can be added (e.g. award type) and more Expenditure_ columns can be added)

2. The database is populated from the data entry spreadsheet and provides the source information for analysis output (see below).

The data entry spreadsheet provides a simple and logical way for data to be captured, but it is difficult to analyse directly and inefficient to store data in this way (where there are large numbers of empty cells). Similarly the analysis spreadsheet necessarily contains repeated redundant data (classification labels).

By contrast the database is "normalised" (it has no empty space or repeated data) and can be used to efficiently store the data over the long term It has a simple three table internal structure linked by the unique identifiers for awards (awards and financial amounts, health categories and percentages, and research activity codes and percentages).

3. The output spreadsheet summarises and analyses the data exported from the database.

It presents the data in several different graphical styles using pivot analysis worksheets. Each worksheet reproduces one of the figures or tables of the two published UK health analysis reports.

Discussion points

2.4.1. Attendees noted that it would be important to keep the HRCSonline website updated, so that it provides a single point of reference for the growing community of HRCS users.

2.4.2. The discussion outlined several areas where a central resource would be helpful:

- Sharing approaches for the validation/quality control of coding
- Keeping the coding up to date – science moves on so there will always be new areas to code for, the HRCS codes and training information therefore need to be kept up to date.
- Training/guidance documentation - tips and pitfalls that are useful for coders. As more classification work is carried out it would be helpful to capture this learning centrally to share.
- Links to published analysis or relevant literature citing the HRCS.

2.4.3. Attendees suggested that support for the community of HRCS users via an online discussion forum may be helpful, and development of the HRCS using a "Wikipedia"-type approach may also have merit.

2.4.4. The discussion also highlighted points made in the previous presentation on the ICRP; could in future there be a central repository of portfolio data coded using the HRCS?

2.5. Classification of research at the DFG (Katarina Fuss, DFG)

The German Research Foundation (DFG,Deutsche Forschungsgemeinschaft) funds research projects in all fields of science and the humanities. This includes support for individual projects and research collaboration, awards for outstanding research achievements, and funding for scientific infrastructure and scientific cooperation. The DFG is the main self-governing organisation for science and research in Germany, and its membership consists of German research universities, non-university research institutions, scientific associations and the Academies of Science and the Humanities.

The DFG receives the large majority of its funds from the states and the Federal Government, which are represented in all Grants Committees. At the same time, the voting system and procedural regulations guarantee that the DFG is independent in its choice of what research to support. In 2009 the DFG spent €2.2 billion (£1.8 billion) on research.

DFG have developed their own coding system which applies across all disciplines and is applied to applications to the DFG. This coding then determines how proposals are allocated to review boards, the choice of reviewers, and how the work is to be evaluated. There are 48 review boards with about 590 members. The DFG research classification system covers 200 subjects, in 14 areas, over 4 main disciplines.

The system is reviewed every four years, so that new developments in science can be accommodated and the operation of the system is optimised.

DFG have examined how their discipline classification can be mapped into the HRCS health categories, with the aim of being able to compare investments with other funding agencies. The work has focussed on those DFG codes in medicine, and initial steps have been to identify parallels between HRCS and DFG coding and whether a concordance table can be drawn up to map codes between the two systems. It seems that 70% fit well, and so it might be necessary to leave the detailed level of categories and to start at an aggregated level of analysis. This work is ongoing.

In addition to the classification by discipline the responsible DFG subject specialist can allocate a further code that refers to topics that are "trans-sector", that cross several sub-disciplines. In Medicine these additional coded themes are;

- human embryonic stem cell research
- cancer research
- cardiovascular research

These codes present some additional challenges in mapping to the HRCS.

The ultimate aim will be to be able to use either classification system to analyse both the input to, and the outputs from research.

Discussion points:

2.5.1. Attendees recognised the merit of a system which organised all areas of research into a common classification, not just the discipline of health research. The use of both coding systems for analysis will add further usefulness and value to the analysis using HRCS alone.

2.5.2. Other funding and research performing organisations can learn from how DFG are translating their own coding system into HRCS health categories. Attendees emphasised the need for approaches to map between classification systems, as it was unlikely that internationally a single coding approach would ever be common among all funding agencies given the wide variation between research organisation missions.

2.5.3. In addition some analyses demanded portfolio data over long time series, and many organisations had evolved, or entirely changed their approaches to classifying their portfolios over time. Methods to map one approach to the other would be needed to join data across time periods.

2.5.4. Attendees also discussed automated approaches to classifying research. Some attendees had experience with commercial text mining approaches, and it was suggested that natural language processing and other computer aided searching of abstracts and/or titles will in future greatly aid the classification of research. Although these approaches currently need careful training, and validation using human intervention, there may be the potential in future to classify and re-classify large numbers of scientific abstracts at low cost.

3.   Summary of Discussion from Workshop Session 1

The discussion can be summarised under four broad areas;

- efforts to ensure consistent classification
- proposals to support wider use of a standard classification system
- methodological suggestions regarding improvements to the efficiency/range of analysis that can be performed
- how portfolio analysis can support strategy development.

3.1.   Consistency of classification

For analysis across research organisations, for example at a national or international level, classification has to be applied consistently to each research portfolio. Assuring this consistency was a major concern for organisations investing in processes to classify their research portfolios. The choice of who should code applications or awards, their training, and processes to double check coding were all important aspects of this quality assurance. Facilitating the exchange of experiences between organisations that analyse their research portfolios was agreed to be vital for any common standard to develop. It is important that best practice can be shared and training materials jointly reviewed and updated. The analysis of 2009/10 health research portfolios in the UK, which is planned in 2011, will have to demonstrate that organisations can implement suitable approaches for quality assurance.

3.2.   Proposals to support wider use of a standard classification system

While the discussion focussed on the use of the HRCS, the workshop was just as relevant to the support of any classification system used across research organisations and in particular internationally. The ICRP, which provided a central repository for coded portfolio data and tools for analysis, was agreed to lead the way in demonstrating the benefits of international collaboration. A central online resource, and forum for organisations actively using the classification system, was agreed as essential. The main components of support are; a single location to access current versions of the classification, guidance and training material, and facility to exchange ideas. There needs to be some oversight of the system by participating organisations in order that amendments/additions to the system can be agreed and introduced. In the UK the Health Research Analysis Forum (HRAF) currently brings together representation from most organisations actively involved in using the HRCS, and reports to the UK Clinical Research Collaboration Board. Materials for guidance on the use of the HRCS are available from www.hrcsonline.net. Consideration should be given to extending the functionality that this site provides or collaborating with the ICRP to provide greater support to the growing community that is using the HRCS.

3.3.   Methodological improvements

The workshop included discussion of the mapping of different research classification systems (e.g. presentation 1.5.). This is a challenging area and methodological improvements are needed in order to more effectively translate codes applied from one system into codes from another.   It was noted that progress had been made in applying text mining technology/natural language processing to abstracts from awards/applications.   These approaches could significantly accelerate the coding process. If successful it may be more efficient to re-code abstracts, rather than translate one classification into another. Methods to quickly re-code abstracts, or translate coding systems would greatly facilitate collaboration between research organisations that wished to compare their portfolios, and would also allow longer time series data to be compiled where previously classification systems had changed or evolved over time.   However automated methods still required human oversight and validation, and so there would be a continuing need to develop collaboratively common standards and approaches for quality assurance.

3.4. Support for strategy development

Most importantly accurate, consistent and comparable information about research portfolios should support work to develop strategy within funding agencies, and ideally co-ordination between funding agencies. Attendees noted that progress had been made in the UK since the introduction of the HRCS by the UK Clinical Research Collaboration with a large number of multi-funder initiatives. Attendees also noted the significant impact that the ICRP was having on collaborative work in the cancer research field across Europe and internationally. In discussion 1.2 attendees noted the impact that analysis using the HRCS was having on strategy development for VR in Sweden. Having consistent portfolio information allows research organisations to determine whether efforts to invest or to de-invest strategically are translated into changes in the portfolio. It also is a powerful tool, particularly when applied to national portfolios to plan to address gaps and new opportunities.

3.5. Next steps

There is a strong connection between the Working Group and the European Medical Research Council who is issuing a Science Policy Briefing on Research Classification before the end of 2011. The results of both groups will be shared to exchange common views and challenges in Research Classification.

Annex 1 – Session 1 Presentation Slides

A1.1 Implementation of the HRCS in Norway (Ingrid Roxrud, The Research Council of Norway)

A1.2 Implementation of the HRCS in Sweden (Jenny Nordquist, Swedish Research Council)

A1.3 International Cancer Research Portfolio (ICRP) (Lynne Davies, Cancer Research UK)

A1.4 Portfolio analysis using the HRCS online tools (Andrew Speakman UKCRC)

A1.5 Classification of research at the DFG (Katarina Fuss, DFG)

All presentation are available on ESF website following the link:
http://www.esf.org/activities/mo-fora/evaluation-of-publicly-funded-research.html

Annex 2 – Acknowledgements

Member Organisation Fora
An ESF Member Organisation Forum is an output oriented, issue-related venue for the Member Organisations, involving other organisations as appropriate, to exchange information and experiences and develop joint actions in science policy. Typical subjects areas discussed in the Fora are related to:
- Joint strategy development and strategic cooperation with regard to research issues of a European nature.
- Development of best practices and exchange of practices on science management, to benefit all European organisations and especially newly established research organisations.
- Harmonisation of coordination by MOs of national programmes and policies in a European context.