Systematic Review Data Repository (SRDR) and Cochrane Data

Prepared by Ruth Foxlee and David Tovey (with thanks to Julian Higgins, Steve MacDonald and Toby Lasserson for helpful comments)

5 November 2011

# Purpose

To gauge the level of support for pursuing data sharing arrangements with the developers of the Systematic Review Data Repository (SRDR), and to identify any specific issues and concerns.

# Urgency

Moderate/high

# Access

Open

# Background

The SRDR is an open-access, web-based, data extraction tool and archive intended for use by producers and users of systematic reviews, developed at the Tufts Evidence-based Practice Center with funding from the U.S. Agency for Healthcare Research and Quality (AHRQ). Deposited data available on the SRDR site could be reused by a wide range of audiences to facilitate the process of subsequent systematic reviews. The underpinning principles of SRDR are the minimization of systematic review production costs and duplication of effort, increased transparency, enhanced data quality (via a process of communal scrutiny) and the fostering of international collaboration.

Current proposals for SRDR suggest that deposited data might include study characteristics, bias assessments and numerical outcome data.

# Proposals and Discussion

The Cochrane Collaboration is seen by the SRDR team as a key partner in their project. David Tovey attended two meetings organised by the SRDR in the US in 2010, as did Kay Dickersin, Karen Robinson and Lesley Stewart (representing CRD). Cochrane Reviews would clearly be a rich source of data for the repository and Cochrane authors an important target group in terms of marketing the SRDR data extraction tool.

Even within study-based registers where key elements of study design are recorded, it is rare for outcome data to be stored, so the first release of the CRS will not serve as a useful data source for SRDR. Future releases of the CRS could contain this type of data, but if it does it will be, in part at least, because we form partnerships with enterprises such as SRDR. In the meanwhile, RevMan is a more appropriate source of data for the SRDR developers. Preliminary discussions have taken place between the SRDR team, the Cochrane Editorial Unit, the IMS team and Metaxis, developers of the Cochrane Register of Studies (CRS). Data-sharing could take three forms:

1. Outcome data from some or all existing Cochrane Reviews could be extracted from RevMan xml files and added to the SRDR data repository as a one-off project or on an ongoing basis.
2. Cochrane Review authors could use the SRDR data extraction tool and deposit these extracted data in the SRDR archive in the process.
3. Cochrane authors and CRGs could use SRDR as a resource which we could be mined for new study data via the CRS.

We suggest that each of these options is considered separately by the CCSG.

Before we pursue any data-sharing arrangements the Collaboration needs to consider the following:

**Benefits**

*Providing data to SRDR*

* By joining SRDR early we could potentially influence their data structures and ensure that linking and data mining processes are optimised to our advantage
* Participating in SRDR signifies the Collaborations' commitment to sharing resources in an effort to raise standards of review production globally and consistent with our recent data access statement

*Using SRDR as a data extraction tool*

* Across the Collaboration multiple data extraction methods are being used. The use of a standard tool like SRDR could lead to more efficient authoring and editorial process

*Using SRDR as a source of data for Cochrane Reviews*

* Data extracted by one team of Cochrane authors would be available to other Cochrane author teams , although we could have this within CRS if we choose to develop the software in this way
* Access to existing data collection forms could promote continuity of methodological approach when new author teams update reviews.
* Extracted data could be scrutinised and quality of Cochrane reviews improved as a result.
* As the SRDR grows unpublished data may be available that would otherwise be excluded from our reviews.
* Linking to SRDR takes us out of our own "data bubble", enabling us to find and use data that have been extracted by other non-Cochrane systematic reviews. The CRS already does this with CT.gov and the SRDR could be used in the same way.

**Challenges**

*Providing data to SRDR*

* SRDR is an open access site so data extracted by Cochrane authors could be reused without proper attribution.
* SRDR might be used by competitors to produce systematic reviews more efficiently irrespective of any issues of reduced quality
* For the resource to be really appropriate, it arguably should have all data available for a study, and not just the data that a Cochrane review team extracts for their particular objectives. There could be important resource implications here for Cochrane authors.

*Using SRDR as a data extraction tool*

* Using the SRDR extraction tool is likely to require training which the Collaboration would not be responsible for delivering, and would not be able to support directly once this was established.
* There are many legitimate differences in the types of data people collect, and a single data extraction template can't work for all. We need to explore how such variation will be addressed if the data set is also to fit into a database structure?
* Only limited piloting of the SRDR data extraction tool has been undertaken by a few Cochrane CRGs. Further investigation of the acceptability of the tool would need to be carried out.
* Other web-based systems are in active development and it could be that collaborating with one system may limit our flexibility to work with other systems.

*Use of SRDR as a source of data for Cochrane Reviews*

* Recent experience with web-based platforms suggests that compatibility between data generated and stored in external systems and Review Manager is still to be optimised. Collaboration between our software developers and developers from external systems could address this issue however.
* Data extraction errors, or misunderstandings, as well as legitimate differences in choices of which outcomes or outcome measures to extract (e.g. 6 months or 9 months?) could cause problems if Cochrane review authors rely on data extracted by other people. We believe that authors will continue to need to examine the study reports in detail in order to produce a review that they can defend publicly.

If the Steering Committee approves the principle of sharing data with SRDR (item #1 on the list above), we propose to carry out further consultation - with CRGs, review authors, methodologists, other Cochrane groups and our publishers to ascertain the level of support for further co-operation with the SRDR programme. We will then return to the Steering Committee with more thorough proposals.

# Summary of recommendations

We recommend that the Steering committee consider the 3 proposed scenarios separately.

1. Outcome data from some or all existing Cochrane Reviews could be extracted from RevMan xml files and added to the SRDR data repository as a one-off project or on an ongoing basis.   
     
   We propose that the CCSC supports further exploration of the feasibility and implementation of this (it should not require anything more than minimal effort on the part of CRGs), but gives approval in principle to such data sharing.
2. Cochrane Review authors could use the SRDR data extraction tool and deposit these extracted data in the SRDR archive in the process.  
     
   The SRDR is only one of many online data extraction tools, and we do not wish to limit review authors to only using the SRDR. However, if "in principle" support is given to option 1, we should explore further the use of SRDR as a data extraction form.
3. Cochrane authors and CRGs could use SRDR as a resource which we could be mined for new study data via the CRS.   
     
   We recognise that there is considerable and understandable concern about this, and the potential for reducing the quality of reviews. Therefore we would wish to discourage this use of the SRDR within Cochrane, pending further exploration of ways to assure good practice. That said, given that SRDR is intended to be an open access resource our ability to "police" such use is necessarily limited.

# Resource implications

None at present

# Impact statement

The consultation on SRDR data sharing could be rolled up within the Cochrane Content preparatory work. David Tovey has been invited to sit on an advisory committee panel for the project, which will take up one hour every alternate month.

There might be very positive impact if Cochrane is seen to be a willing partner of this project in terms of supporting the availability of data from trials and also for facilitating a more substantial relationship with the Tufts EPC, and AHRQ.

There are also potential benefits in terms of:

* Reducing the risk of data loss by storing and archiving data.
* identifying data extraction errors
* more efficient data extraction
* using data from the SRDR to promote more efficient systematic review production.

Harmful impacts include the risk that Cochrane authors will rely inappropriately on un-checked deposited data leading to quality problems

# Decision required

The Steering Committee is asked to consider whether pursuit of data-sharing with SRDR would be potentially of benefit to the Collaboration and merits further exploration. If the approval of this course of action is received a proper evaluation of the resource and other implications will be undertaken. The Steering committee might also identify additional priority issues that should be part of any further exploration.