HarmoniSR: final project report

September 2015

Current members of the HarmoniSR group:

Anna Noel-Storr (TSC Exec); Liz Stovold (TSC Exec); Liz Doney (TSC Support team); Sam Faulkner (TSC Support Team & TSC Exec); Ruth Foxlee (Cochrane Editorial Unit); Gordon Dooley (Metaxis)

Background

The HarmoniSR group was originally formed to try and standardise the format of trial registry records in the CRS. It quickly became apparent that this work should also encompass mappings for CRS study records, as well as other types of reference records where there is currently variation in mapping and content of fields.

The aims of the project were:

1. To develop guidance for the formatting of reference records in CRS for four reference types: Journal article; clinical trials registry record; conference abstract in a journal; conference abstract in conference proceedings.

2. To develop a core set of study fields, consistent with the fields in the Linked Data PICO annotator, which we can recommend people use when populating their study records in the CRS.

Methods

Prior to the current and final stage of this work, three rounds of consultation had taken place on the ideal mapping of trial registry records, conference records and journal records. There were webinars, presentations, and workshops at Cochrane meetings looking at both reference and study records.

In March this year, we circulated two documents: one was regarding reference records, with suggested mappings for four reference types; and one regarding study records, with a set of core study fields. We asked for comment and feedback from the TSC community on these documents.

In June this year we secured some funding through Cochrane's discretionary funding scheme to complete this work. We collated all the feedback we received on both reference and study records and worked through each field one by one for each record type during a

series of six conference calls and one face-to-face meeting. For each field in each record type, we considered the feedback, any technical issues or solutions, made a decision on each field by consensus, and recorded our decision, together with some suggestions for guidance.

Results

We received feedback from 6/52 (11%) on the reference document, and 7/52 (13.5%) on the study document. Through discussion we were able to agree on the core fields for the four reference type records and for a study type record. For reference records we considered how the information stored in a CRS record could be displayed in the published CENTRAL version and exported to RevMan. During our discussion about the study record, we took into account the fields included in the Cochrane PICO annotator tool, and the fields available to us in a ClinicalTrials.gov record.

During the discussion process we generated a list of additional required reference fields for both CRS Web and CENTRAL (see below). A request to Wiley will be made for additional publishable fields in CENTRAL. In the interim, existing fields will be used so that all pertinent information needed to retrieve the citation will be published in CENTRAL.

Reference records in the CRS

Full guidance for the format of the content for each field is expanded upon in Appendix A. The core fields for each reference type are as follows:

1) Journal article

Field	CRS Field Tag	Already in CRS?	CENTRAL field?
Author	AU	Yes	Yes
Title	TI	Yes	Yes
Original title	OT	Yes	Yes
Source	SO	Yes	Yes
Year of publication	YR	Yes	Yes
Volume	VL	Yes	Yes
Issue	NO	Yes	Yes
Pages	PG	Yes	Yes
Abstract	AB	Yes	Yes
PubMed accession number	PM	Yes	Yes
Embase ID accession number	EM	Yes	Yes
DOI	DO	Yes	Yes
MeSH headings	MH	Yes	Yes
MEDLINE publication type	PT	Yes	Yes
URL of full text	UF	Yes	Yes

Owner	CC	Yes	Yes
Language published	LA	Yes	Yes
Trial registration number	TN	Yes	No
Other database ID number	ID	Yes	No
Study design	DE	Yes	Yes
Emtree headings	EMT	Yes	Yes
Pubmed Central ID	PMC	No	No
Article Identifier (pii)	AID	No	No
Article ID (for Epubs)	EI	No	No
Publication Status (currently called 'status of report)	ST	Yes	No

2) Conference abstract in a journal

We came to a major decision on this record type. In response to feedback provided during the latest round of consultation we re-thought our proposal to try and parse conference details out into separate fields. For conference publications published within a *journal*, the most important thing is that journal details are entered accurately and consistently. The conference details will all go into a single field in CRS (a new conference details fields). Those details can be published to CENTRAL in the Comments field, which is a publishable field. In the long-term we will request a new field for conference details in CENTRAL. The rational for our decision: 1. we are putting our efforts into making sure the citation details are consistently reported; 2. conference details are very hard to parse accurately and we don't want to generate a lot of manual editing.

There is currently variation in practice around where to put the abstract number. Some include it in the page field; others in the title field. We made the decision to keep individual pieces of bibliographic information separate to allow for flexibility; therefore a new field will be introduced in CRS Web for abstract number. We will request a new field in CENTRAL, but until then, this field will append to the page field when exported to CENTRAL format so it is included in the CENTRAL record.

Field	**CRS Field Tag**	**Already in CRS?**	**CENTRAL field?**
Author	AU	Yes	Yes
Title	TI	Yes	Yes

Original title	OT	Yes	Yes
Source	SO	Yes	Yes
Year of publication	YR	Yes	Yes
Volume	VL	Yes	Yes
Issue	NO	Yes	Yes
Pages	PG	Yes	Yes
Abstract number	AI	No	No
Abstract	AB	Yes	Yes
Embase ID accession number	EM	Yes	Yes
DOI	DO	Yes	Yes
URL of full text	UF	Yes	Yes
Language published	LA	Yes	Yes
Trial registration number	TN	Yes	No
Other database ID numbers	ID	Yes	No
Study design	DE	Yes	Yes
Emtree headings	EMT	Yes	Yes
Conference details	CD	No	No
Proceedings location	**PL**	**REMOVE FROM CORE FIELDS***	
Proceedings date	**PD**	**REMOVE FROM CORE FIELDS***	

*If Groups wish to continue using these fields, they can be set up as private fields

3) Conference abstract in a conference proceedings

For this record type, conference details should go in the source field, as they are formatted in the source record, and should not be edited to include 'Proceedings of'. We will not attempt to parse those details into individual fields. Not all of these fields will be present in this publication type (e.g. Volume, issue) but they are listed here for completeness.

Field	**CRS Field Tag**	**Already in CRS?**	**CENTRAL field?**
Author	AU	Yes	Yes
Title	TI	Yes	Yes
Original title	OT	Yes	Yes
Source (conference details)	SO	Yes	Yes
Year of publication	YR	Yes	Yes
Volume	VL	Yes	Yes

Issue	NO	Yes	Yes
Pages	PG	Yes	Yes
Abstract number	AI	No	No
Abstract	AB	Yes	Yes
Embase ID accession number	EM	Yes	Yes
DOI	DO	Yes	Yes
URL of full text	UF	Yes	Yes
Language published	LA	Yes	Yes
Trial registration number	TN	Yes	No
Other database ID numbers	ID	Yes	No
Study design	DE	Yes	Yes
Emtree headings	EMT	Yes	Yes
Proceedings location	**PL**	**REMOVE FROM CORE FIELDS***	
Proceedings date	**PD**	**REMOVE FROM CORE FIELDS***	

*If Groups wish to continue using these fields, they can be set up as private fields

4) Trial registry record

This record type has been the source of the greatest amount of uncertainty with regard to formatting. We have had to be pragmatic and make decisions that will allow all the pertinent information needed to identify the trial record from a CENTRAL citation:

- Until we have a TN (trial registration number) field published in CENTRAL, the author field in a trial registry record will contain the trial registration number. We recognise that this will duplicate information held in the TN field in CRS.
- The source field will contain the full URL to the record.
- There will be a new field in CRS for Trial Registry Name. This will be used to record which registry the record was identified from (e.g ClinicalTrials.gov; WHO ICTRP)
- The Trial registry name (TR) field and the SO field will combine on export to CENTRAL so that both the registry name, and full URL, are present in the CENTRAL SO field. We recognise that there will be duplication of the URL in both the SO field and UF field.
- Trial registry records have two titles, a public title, which will go into TI, and an official/scientific title, which will go into OT. We recognise that this is a different

use of the OT field from other record types but it is an alternate title, therefore this is an appropriate use of the field.

- The YR field will be used for 'date first received' and will just contain the year.

Field	CRS Field Tag	Already in CRS?	CENTRAL field?
Author	AU	Yes	yes
Title	TI	Yes	yes
Original title	OT	Yes	yes
Source	SO	Yes	yes
Year of publication	YR	Yes	yes
Abstract	AB	Yes	yes
Trial registration number	TN	Yes	No
Study design	DE	Yes	Yes
Trial Registry name	TR	No	No
URL of full text	UF	Yes	Yes
MeSH headings	MH	Yes	Yes

Proposed new fields for CENTRAL

The current publishable CENTRAL fields are listed in Appendix B. We are proposing, as a longer term project, that the following fields be added for publication in CENTRAL:

Field	Tag	Already in CRS?
Record type	RT	Yes
Date published	DP	Yes
Abstract number	AI	No
Conference details	CD	No
Trial registration number	TN	Yes
Trial registry name	TR	No
PubMed Central ID	PMC	No

Study records in the CRS

The aim of this part of the project was to agree a "core" set of fields for a study record in CRS. This does not imply that all Review Groups must populate their CRS study records, but for those that already do, or for those Groups who wish to begin populating study records with meta-data, these are the main fields to describe a study in the CRS. Review Groups may use as many other fields as they like in their study records. These fields are the minimum set to be used consistently across Groups.

For all core fields where it might not be possible to complete them as the information is not available, there will be a radio button to select 'not applicable' so that no core fields

are left blank. Records and fields that are auto-populated from CT.gov will update automatically when the source records are updated.

Each new CRS study record created, either manually or automatically, will be given the status of 'pending' in the SSI field.

Guidance for the format of the content each field is expanded upon in Appendix C. The core fields for a CRS Study Record are:

Field	Field tag	Already in CRS?
CRS study ID		Yes
Study name	SN	Yes
Study full name	SFN	Yes
Study acronym	SA	Yes
Study Title (Public)	SCT	No
Study Title (Official)	STI	No
Study registration ID	SID	Yes
Other study IDs	SON	Yes
URL to study registration	SUR	Yes
Study design	SDE	Yes
Healthcare condition	SCO	Yes
ClinicalTrials.gov health care condition	SCC	No
Intervention type	SIT	No
Interventions	SIN	Yes
ClinicalTrials.gov intervention	SCI	No
Participants (number)	SSZ	Yes
Study Outcomes	SOC	No
Expected end date of study	SXD	Yes
Data extraction status(currently called status of study information)	SSI	Yes

The following fields were discussed, agreed as non-core fields, and so removed from the core field list, but will remain available fields:

Field	Field tag	Comments
Status of study	SS	Can be updated automatically from CT.gov
Treatment setting	STS	Not currently included in PICO tool
Primary outcomes	SPO	Can be auto-populated from CT.gov, or used to store

		legacy data. The core outcome field is 'Outcomes'.
Secondary outcomes	SSO	As above.

The following fields were proposed during the feedback period. We discussed and agreed they are non-core fields which won't be included at this stage, but may be reviewed at a later date. CRGs who wish to use these fields are free to create their own private fields.

Field	**Field tag**	**Comments**
Dose/duration etc		In PICO annotator tool.
Outcome timepoints		
Language		
Country of origin		
Reason for exclusion from review		

Discussion

The reference and study fields discussed in this report are the recommended "core" fields. Groups are free to use as many additional fields as needed.

We recognise that some fields in reference records actually contain study information (e.g. study design; trial registration number). While CENTRAL remains a database of references, these fields need to be included in a CRS reference record so they can be published to CENTRAL.

We are incorporating many fields available to us in CT.gov with the aim of enriching our data. We have also taken into account fields in the Linked Data PICO annotator which correspond with a CRS study record.

As explained in the section on study records, populating study records with meta-data in a Group's specialised register in the CRS is not being made mandatory. The guidance on study records is for those that already do this, or want to begin this activity.

Much of the feedback we received during the consultation process was around CRS records differing from the format required by the Style Guide and RevMan. Where possible, records in CRS Web will be automatically converted on export to the required format, for example, the page field will be converted from CRS format (e.g. 325-327) to RevMan format (e.g. 325-7) when exported to RevMan. In addition to these technical solutions, the HarmoniSR group has had the opportunity to provide input in to the revision of the reference section of the Cochrane Style Guide, and we have made some

suggestions which we hope will bring CRS and RevMan formatting into closer alignment in future.

Implementation of HarmoniSR guidance will not formally commence until the roll-out of CRS Web, although TSCs are welcome to begin using the guidance as they populate their registers in the meantime.

Next steps

- Guidance and core fields built into CRS Web
- Standalone, printable guide to populating core fields to be generated
- Officially 'retire' the 'TSC User Guide to managing specialised Registers and handsearch records'.
- Set up a HarmoniSR section on the TSC portal where this report and other related documents will be stored.
- ClinicalTrials.gov will be the pilot database for the Centralised Search Service.
- Evaluating/editing default CRS import filters to comply with new guidance
- Evaluating/editing default CRS export options to comply with CENTRAL & RevMan requirements
- Request new publishable fields to be added to CENTRAL records
- Background standardisation changes to be made where possible by Metaxis

Appendix A

CRS core field guide – for reference records

NB CENTRAL fields are in pink.

Proposed fields will be available in CRS Web. Creating them in CENTRAL will be longer term.

Abstract (AB)

For journal records reproduce the abstract as it appears in the original record. For copyright reasons some abstracts will be removed when records are submitted to CENTRAL, though they are retained in your segment.

For trials records use the purpose/description information.

Abstract number field (AI)

(Proposed). To be alpha numeric. To be appended to the page field in CENTRAL until a field can be added.

Article ID numbers (EI)

(Proposed). For ePubs.

Author (AU)

For journal records use last name followed by initial/s. Authors separated by // e.g. Seboka G // Saunderson P // Currie H Jr. Leave field blank if authors are not known - do not use Anon.

For trials records the trials registration number currently maps to the author field. It should also map to the TN field.

Conference details (CD)

(Proposed). Reproduce as it appears in the original record. Will be published to the comments field in CENTRAL until a field can be added.

Date published (DP)

For trials records use the ‘date first received’ in full e.g. 2015 May 23.

DOI (DO)

Enter the DOI, removing http:// or https:// Ensure this field is populated for ePubs. DOIs always start with 10 e.g. 10.1038/ajg.2015.191.

Embase accession number (EM)

Enter the Embase number, do not prefix with 'EMBASE'. Do not add any other database IDs.

EMTREE headings (EMT)

Use // to separate EMTREE terms. Use / to separate a term from it's qualifier e.g. Hand Dermatoses/diagnosis.

Issue (NO)

Number of the journal issue in which the article was published. Do not use 'Iss'. Permitted abbreviations:

Part – Pt

Supplement - Suppl

Special Number – Spec No

Special Issue – Spec Iss

Anniversary number – Anniv No

Language published (LA)

Use the pick list of ISO codes. First letter is not capitalised. You can select more than one language.

MEDLINE publication type (PT)

Use // to separate the terms.

MeSH check words (MC)

Use // to separate the terms.

MeSH headings (MH)

Use // to separate MeSH terms. Use / to separate a term from it's qualifier e.g. Hand Dermatoses/diagnosis.

Original title (OT)

Use only for non-English language titles. Use sentence case, and no full stop at the end. Letter after a colon is lower case. No square brackets. No language label.

For trials records use the scientific title.

Other ID numbers (ID)

Use for IDs for references from databases other than PubMed and Embase, e.g. CINAHL, PsycINFO. Use // to separate IDs.

Owner (CC)

Automatically populated when a record is published in CENTRAL.

Pages (PG)

Use full pages format with a hyphen and no spaces e.g. 33-39. For ePubs use an enumber if there is one e.g. e0130145, and put a DOI in DOI field. Do not add text such as 'ePub'. Do not put abstract or poster numbers here, use the Abstract number field. For discontinuous page numbers write the ranges separated by a comma, e.g. 3, 6-8.

The page field will be converted on export to RevMan format (e.g. 33-39 will become 33-9) to comply with Cochrane Style Guide.

PII (AID article identifier field)

(Proposed). Number only. Do not prefix PII.

PMC number

(Proposed). Number only. Do not prefix PMC

Publication status (PST)

(Proposed change of name from Status of report (ST)). Use for ePub ahead of print.

PubMed accession number (PM)

Enter the PubMed number, do not prefix with PMID. Do not use for any other IDs, e.g. PMC.

Record type (RT)

(Proposed change of name from Type of report (RT)). To be a picklist: Journal article; Conference publication; Trial record; Letter; Other.

Source (SO)

Select from NLM authority list if possible. Use sentence case. Use full journal name. Only include the English translation in square brackets after a foreign language title if a translation is provided by the original journal or database.

For trials records, type the URL in full including http:// or https:// to the trial record (published to CENTRAL alongside the trial registry name).

For conference proceedings, enter the name of the conference as it appears, do not add 'Proceedings of' if this is not included.

The source field will be converted from sentence case to title case on export to RevMan format to comply with Cochrane Style Guide.

Study design (DE)

Use the pick list of RCT // CCT // ITS // CBA. If you are unsure if a study is an RCT, use CCT. If you keep other kinds of reference in your register e.g. systematic reviews or meta-analyses, create a separate, private, field to classify these publication types.

Title (TI)

Use sentence case with no full stop at the end. Letter after a colon is lower case. No square brackets. Do not label with publication type e.g. Abstract.

For trials records use the Plain language trial name.

Trial registry name (TR)

(Proposed). Enter the source trials register e.g. ClinicalTrials.gov. To be published to the source field in CENTRAL (alongside the URL from the source field) until a field can be added.

Trials registration number (TN)

Use the study registration identifier with no spaces e.g. NCT01132651. Use // to separate several identifiers for the same trial. Example: NCT01083433 // ISRCTN16250774.

URL

If an abstract or full-text paper is freely available, enter the URL here. This field is published in CENTRAL. It should not contain links to documents that are not freely available. Type the URL in full including http:// or https://.

For trials records use the URL to the trial record.

Volume (VL)

Number of the journal volume in which an article is published. Do not use 'Vol'. Permitted abbreviations:

Part – Pt

Supplement - Suppl

Special Number – Spec No

Special Issue – Spec Iss

Anniversary number – Anniv No

Year of publication (YR)

Enter the four digit year e.g. 2015. Use the DP field for fuller date of publication including day or month.

For trials records use the 'date first received' and just enter the year.

For conference proceedings, the year of the conference is not always the same as the year of proceedings publication; enter the year of proceedings publication here.

Appendix B: current published fields in CENTRAL

Field
Cochrane group code
CENTRAL ID
Title
Authors
Source
Year
Volume
Issue
Pages
Abstract
MeSH
MeSH check words
Correspondence address
Pubmed ID
Embase ID
DOI
Medline publication type
Original language title
Editors
Publisher
Publication date
Place of publication
Study design
Language
URL of summary or abstract
URL of full text article
Comments
CRG keywords
EMBASE keywords

Appendix C

CRS core field guide for study records

Fields high-lighted in blue can be auto-populated or auto-generated and do not require any manual data entry or editing.

ClinicalTrials.gov healthcare condition (SCC)

Auto-populated from trial registry source record.

ClinicalTrials.gov intervention (SCI)

Auto-populated from trial registry source record.

CRS study ID

Auto-generated unique ID number.

This field allows for any variation in the SN field across reviews and CRGs.

Data extraction status (SSI)

Status defaults to 'Pending' when record is created.

Use pick-list to update status: in progress//complete

Expected end date (SXD)

Auto-populated from trial registry source record, or entered manually.

If entered manually the date format should follow the convention: Year Month (abbreviated).

Select ‘not applicable’ for completed trials.

Healthcare condition (SCO)

Select condition(s) from pick list. This field uses the same vocabulary as the Linked Data PICO annotator tool.

Annotate all conditions, not just condition of CRG interest. Multiple terms allowed and will display in in CRS as delivered by PICO tool.

Intervention type (SIT)

Select from pick-list.

This field uses the same classification list as the Linked Data PICO annotator tool.

Interventions (SIN)

Select intervention(s) from pick list.

This field uses the same vocabulary as the Linked Data PICO annotator tool. Multiple terms allowed and will display in CRS as delivered by PICO tool.

Other study IDs (SON)

Auto-populated from the trial registry source record or entered manually.

This could be a drug company ID or funders ID. This field is not for trial registration numbers. Multiple numbers are allowed. Separate multiple numbers with '//'.

Study design (SDE)

Select from pick-list: RCT // CCT // ITS // CBA

If coding other study designs then use a private field to record the study design

Study full name (SFN)

Entered manually.

This is the study's' given name or expanded acronym. This is not the full scientific title.

Select ‘not applicable’ if information is not available.

Study name (SN)

Auto-generated using Cochrane’s naming convention for studies (Surname Year) or entered manually.

If entering manually you may follow your CRG convention, but bear in mind this is the field that will be exported to RevMan and should therefore comply with the Cochrane style guide.

The study name may change over time as the study moves from registration to completion and results are published.

Study registration ID (SID)

Auto-populated from the trial registry source record or entered manually.

Both international and national trial registration numbers are allowed. Multiple registration numbers allowed. Separate multiple numbers with ‘//’.

This field is not for drug company trial numbers which should go into the ‘Other study IDs’ field.

If you can’t find a registration number, select ‘not identified’.

If you select 'not registered’ then you must have made substantial efforts to check published papers and trial registries, and you may wish to record this in a notes field.

Study title (Official) (STI)

Auto-populated from the trial registry source record.

This is the full scientific title.

Study title (Public) (SCT)

Auto-populated from the trial registry source record.

This is the Public/lay title.

URL to study registration (SUR)

Auto-populated from trial registry source record or entered manually.

If entering manually, type full URL to the trial registry record including http:// or https:// Multiple URLs are allowed. Separate multiple URLs with <space>//<space>