



---

# #CochraneTech to 2020: The role of linked data in meeting our strategic goals

---

Draft presented to the Cochrane Collaboration Steering Group  
July 2013 – **Financials redacted**

by the Cochrane Linked Data Project Board

Co-chairs:

Chris Mavergames, Director of Web Development  
Jessica Thomas, IMS Team Manager

Project Board:

Lorne Becker, Cochrane Innovations  
Chris Champion, Managing Editor, CCDAN, & ADAC convenor  
Gordon Dooley, Metaxis/CRS  
Julian Elliott, Monash University  
Toby Lasserson, Senior Editor (CEU) & RAC Co-Convenor  
Rasmus Moustgaard, Senior Developer, IMS Team  
Juliane Ried, Project Officer, Cochrane Innovations  
Karen Robinson, Johns Hopkins University  
Sheila Wallace, TSC rep. on RAC



## Table of Contents

---

1. What's this all about and why should we consider doing it? .....	4
1.1 Background and framework .....	4
1.2 Why consider a change? .....	4
Open Access .....	4
The Cochrane Author Experience .....	5
Open Data .....	5
1.3 Brief background on Linked Data and Semantic Web .....	5
1.4 Building on what we have.....	6
1.5 What happens if we don't do it / Alternative options.....	6
2. What are we proposing? .....	7
2.1 Over-arching summary .....	7
2.2. Key recommendations .....	7
Create a linked data store to classify and organise Cochrane content and to tie together existing Cochrane databases and software systems.....	7
Build Cochrane ontologies that defines key components of Cochrane content and their relationships with one another.....	8
Identify appropriate vocabularies or taxonomies from controlled and standardized sources to use in tagging our content with metadata.....	9
Adopt a Service Oriented Architecture to improve inter-system communication.....	10
Employ 'user stories' to identify and prioritise potential software development activities.....	11
3. How will we benefit? .....	12
3.1 Cochrane Collaboration .....	12
Systematic Review Ontology.....	12
Better delivery of content and interaction with third parties .....	13
Sustainability.....	13
3.2 Entity staff.....	13
3.3 Authors .....	13
3.4 Users of Cochrane content .....	14
4. Challenges of the project.....	14
4.1 Organisational and cultural challenges.....	14
4.2 Technology teams and our Publisher .....	14
4.3 New ways of tagging and displaying our content and data might bring concerns.....	14
4.4 Methodological issues need to be dealt with.....	15
4.5 Managing expectations.....	15
4.6 Evaluation - "How will we know we've been successful?" .....	15
5. What's the larger context? .....	15
6. How will we proceed? .....	16
6.1 Gradual iterative approach.....	16

Foundation Stage - Establishing basic linked data capabilities.....	17
Hire or contract with key personnel .....	17
Buy or license the needed hardware & software .....	17
Begin work on core Cochrane linked data structures.....	18
Strengthen existing Cochrane teams and build partnerships with collaborators .....	18
Exploration Phase - Using linked data to address selected user stories .....	19
Production phase .....	19
Importance of evaluation at every stage.....	19
8. Summary.....	19
Our principles.....	20
Our strategic goals .....	22
Building on the Information Services Strategy Framework and Structure.....	22
Appendix.....	24
Related documents – Attached as separate documents.....	24
Linked Data Background paper: From December 2012 meeting in London .....	24
Ontoba report: <i>Linked Data at The Cochrane Collaboration: A technical strategy</i> .....	24
Figure 1 - Linked Data Service-oriented Architecture: Diagram from the Ontoba Report.....	25
Figure 2 – Cochrane Review prototype ontology from the demonstrator.....	26

## 1. What's this all about and why should we consider doing it?

---

### 1.1 Background and framework

---

This paper recommends that the Collaboration funds and supports adoption of a linked data approach to our technology infrastructure. Adopting this approach will enhance the sustainability of our organisation by improving efficiency in the development of our software and the production of Cochrane Reviews, and enhancing visibility and flexibility of our content. It is important to note that this project is as much about our content as it is about our technology.

This change is not seen as an abrupt replacement of current structures and development processes. Proposed is a gradual, planned evolution that maintains the considerable strengths of what the Collaboration's software teams have already developed, but adapts their processes and Cochrane data to this new approach.

Some important background to this recommendation is included in a Cochrane Linked Data Background Paper prepared last November. That paper includes a brief description of linked data and some ideas about its potential usefulness for the Collaboration. It also describes the Cochrane linked data pilot project that resulted in a demonstrator that was used to inform the Collaboration's strategic discussion in Paris. Readers of this paper who are unfamiliar with the concepts of linked data may wish to read that background paper before reading this one. It is attached here as a separate document.

Throughout the linked data pilot project, we were guided by a team of consultants from Ontoba – a digital media consultancy group with expertise and specialist knowledge in semantic web software and solutions architecture. At the end of the pilot phase, our Ontoba consultants prepared a report containing specific recommendations for the Collaboration in moving to a linked data approach. That document is available for review as a separate document as well, but much of it is quite technical. The Cochrane Linked Data Project Board reviewed the document in detail and discussed each of the recommendations. We have attempted in this paper to summarize what we see as the key issues from the Ontoba document, to outline benefits to The Collaboration arising from adoption of these changes, and to give some idea of the sort of investment needed.

The proposed approach represents an attempt to “future proof” Cochrane in order to make us more sustainable in the way we produce and deliver our content. The proposed changes are on our side and need not affect our current publisher, Wiley. They are aware of this work being in the pipeline, and have supported the early development, and the plan is to engage them along the way so that they can consume and publish our content and data in new, and hopefully revenue-generating, ways.

A promising, recent development is that Wiley have indicated they will move to a Content Management System (CMS) for publishing *The Cochrane Library*. This system, dotCMS, already has capabilities in the area of linked data, and there are plans to add further features. The proposed changes will not impact regular publication as it currently occurs, but we'll be positioning ourselves to be ready when Wiley are able to work with these new technologies. The bottom line is that Cochrane will be in a position to deliver our content to any publisher, to deliver data to our end-users, and to deliver content and data via new mediums such as Application Programming Interfaces (APIs), in a way that avoids “vendor lock-in” with any specific publisher.

### 1.2 Why consider a change?

---

Our current approach to creating and presenting our content has served us well, but a number of factors in today's environment suggest that it is time for a change.

[Open Access](#)

---

Our new publishing contract calls for a move to open access. This will require us to rapidly explore new models for delivery of Cochrane content - with the possibility that the same content will be used in multiple ways for and by different end-users. Our software development processes will need to be flexible in order to make our content nimble<sup>1</sup>. We can expect multiple test runs of potential derivative products with resulting calls to reconfigure the presentation of Cochrane content in new ways and with new interrelationships.

### The Cochrane Author Experience

Authoring and editing of Cochrane Reviews is difficult and time intensive. As the Collaboration adds new review types, methodological standards and increasingly complex analyses, we need to find ways to make our processes easier and more fluid. The Auckland Colloquium included a number of presentations of new technologies that might help - software screening of citations to identify candidates for included studies, crowdsourcing of some parts of the review process, and software developed outside of the Cochrane IT process to facilitate collaborative review production by author teams. Unfortunately, none of these innovations can easily be incorporated into our RevMan-centric authoring process. Even potentially helpful tasks that use existing Cochrane content (such as importing trial data or Risk of Bias assessments extracted by a different author team) currently require extensive cutting and pasting by authors because of a lack of an agreed data exchange format.

### Open Data

Initiatives such as AllTrials and the Systematic Review Data Repository are part of a broader societal push for increasing open access not only to research reports but also to research data. Cochrane review production software needs to be able to take advantage of such data repositories and to provide data to them when appropriate.

### **1.3 Brief background on Linked Data and Semantic Web**

*“Cochrane needs to get better at talking to machines.” – Ben Goldacre, UK Symposium, 2013*

In brief, linked data can be thought of as supporting the creation of a 'neuro-network for computers' by creating links, encoded in machine readable form, to give the web and computers 'intelligence' on where to identify or how to recognise data. Ultimately this can mean that the data are much more flexible as it allows for the possibility of creating information we haven't yet recognised we need. This has the potential to add enormous value and decrease the need for specific programming as part of large projects.

The goal of the linked data approach is to make finding, sharing, and combining information easier. In the decades between its conception and its realisation there have been many alternative attempts to achieve that goal outside the linked data paradigm using a variety of tools and techniques, but none of these has fully succeeded. The problem has always been with data structures that without knowing how other people's data are structured you can't begin to make your data compatible or to create new structures or products that combine data from multiple providers.

A case in point (with direct parallels to Cochrane data) is Digital Resource Management (DRM) software used by libraries to catalogue their holdings. Almost since the start of the Internet, libraries have looked for ways to make their catalogues searchable online, and a plethora of DRM software has been written over the years that tried to address that aim. Today there are a handful of DRM packages in use by major academic institutions and government libraries which until recently have

---

<sup>1</sup> An adjective in its own right, but see also *Nimble*, a white paper on the future of digital publishing: <http://nimble.razorfish.com>

been largely incompatible with each other. The consequence was that to find an item within a library was relatively easy, but to search *across* library catalogues was largely impossible. In the last few years, all of the main DRM packages have adopted a linked data approach to storage and retrieval of their catalogues with the result that now most library catalogues can be accessed and interrogated by machines, and links between catalogues can be followed quickly and automatically. The first consequence of that is that searchers find more easily the items they are looking for in diverse repositories. More important though, searchers can find things they may not have found before because they can interrogate catalogues by the knowledge they contain rather than by the simple keyword search of legacy systems. DRM software that does not comply with linked data standards is rapidly becoming obsolete.

#### **1.4 Building on what we have**

---

Much of our past history and our current approaches to review production put the Collaboration in a good position to move to a linked data approach. Our "journal" has been structured as a database from its inception. While most of the linked data world struggles with ways to convert text documents into more machine-friendly formats, our Reviews are already in a structured XML format and housed within a data repository that handles versioning and other complexities of the current authoring processes. In addition, the structure of the Cochrane Register of Studies (CRS) and its linkages with the Cochrane Database of Systematic Reviews (CDSR) have given us a model of the literature that can appropriately incorporate trial reports, studies and Reviews, and provides us with a unique ability to navigate between them. The move to linked data would build on these strengths and link them to relevant systems and data repositories developed by others.

The limited moves we have made to explore linked data to date have led to enthusiastic expressions of interest from a number of potential collaborators who recognise the significant strengths of the Collaboration and our major potential role if we adopt this approach. To date, we have been asked to collaborate in:

- a grant proposal to the US National Library of Medicine on linked data approaches to drugs and drug interactions;
- a consortium with representatives from GIN, clinicaltrials.gov, GRADE/DECIDE and others to discuss ways to structure our data so that linkages could be made between trials, trial reports, Cochrane Reviews, guidelines and electronic health records;
- a half day discussion at a major semantic web conference of systematic reviews and their implications for restructuring the article in a linked data world.

#### **1.5 What happens if we don't do it / Alternative options**

---

During our exploration of linked data, no strong alternative has emerged. In fact, it is hard to imagine an alternative emerging that was not interoperable with linked data. Others agree: the BBC, the New York Times, the Library of Congress, the US Department of Defence, the British Museum, Elsevier, Oracle, IBM, Adobe, Yahoo! and Google are all now moving their systems to linked data. For some of those organisations, the move from traditional data storage technology to linked data represents a very large financial commitment – an important indicator that linked data technology has reached a mature stage of development and a clear signal that key players are buying into the linked data approach wholesale. It is the nature of technology that something better may well come along at some point, but when, or if, that happens, the fact that the linked-data approach has already been embraced by many large and influential organisations will mean that a smooth transition from linked data to any new process will be part of the transition process and will likely be straightforward.

The choice for all major content providers is not whether to adopt a linked data approach but when. Asking "do we need linked data?" today may be analogous to asking "do we really need a website?" 15 years ago. If the Collaboration chooses to stick with its existing model of distinct, unconnected

databases, we believe it will miss a unique opportunity to be at the forefront of this revolution, and miss the opportunity to lead the way by setting the standards that others will follow.

## 2. What are we proposing?

---

### 2.1 Over-arching summary

---

Our proposal involves more than simply a shift to a new technology. It will require a fundamental change in the way Cochrane data are conceptualised, stored and used, and in the Collaboration's approach to on-going software and product development. Much of the work to be done in moving to the new approach cannot be accomplished by the Collaboration's software teams working in isolation, but will require active involvement from key Cochrane groups and contributors. Again, this project is as much about content as it is about technology.

A fundamental difference between the linked data approach and current Cochrane processes is the decoupling or "freeing" of the various pieces of Cochrane content (i.e. our essential **data**) from the **user interfaces** that have previously been used to generate them or access Cochrane content. Thus, the various Cochrane IT teams would not have sole responsibility for creating the user interfaces that would access their particular pieces of Cochrane data, but would be responsible for designing software gateways (APIs) that would allow software designed by others (within or outside of the Collaboration) to access those data in a controlled fashion.

The recommendations also have implications for the ways Cochrane contributors think about requests for changes in the capabilities and functions of Cochrane software. Our current approach asks for users to think about changes within the constraints of a given piece of software and add their request to the relevant wish list (Archie, RevMan, CRS etc.). The new approach will call for users to describe what they would like to be able to accomplish and why (through "user stories") without necessarily making reference to a specific piece of Cochrane software.

These new approaches, agile development and user-centred design, have become standard practice in the wider web and software development world. Agile development aims to "end the cycle of endless feature lists" and move from top-down "waterfall" development that features big, but infrequent, software releases to an iterative approach that is more responsive to users' needs. User-centred design requires iteration at its heart – design, test with users, refine, test with users again, refine and then repeat until it is right. Taken together, the agile development paradigm and the user-centred design approach can work together to make Cochrane more nimble in how we both produce and deliver our content.

### 2.2. Key recommendations

---

#### [Create a linked data store to classify and organise Cochrane content and to tie together existing Cochrane databases and software systems](#)

A linked data store (also called a 'triple store') is one of the essential components of a linked data approach. This would be a new data store that would have a Cochrane-wide scope, rather than relating to only a portion of our data. It would not replace the existing data stores (such as Archie and the CRS). These would remain and would be the "truth masters". Much of the content of the new linked data store would be metadata (i.e. data that describes data) that would point to and describe discrete pieces of Cochrane data in the truth master data stores. It would also contain the necessary ontology models (see later recommendation) in a format that would allow searches or queries of the entire repository of Cochrane content and data using standard tools such as SPARQL (Simple Protocol and RDF Query Language), the query language for the Resource Description

Framework (RDF), the data model for linked data. Thus, the linked data store (or triple store) would co-exist with Cochrane's existing subsystems and provide a semantic index of all of Cochrane's data. This approach was used to build the demonstrator for the Paris Mid-Year meeting. Our consultants from Ontoba worked with us to design and build a linked data store that tied together specific pieces of data from CRS, Archie, and external linked databases such as SNOMED CT and Drugbank (<http://drugbank.ca>).

### Benefits to the Collaboration

---

- Improve interoperability between Cochrane's various, existing systems and software. For example, because the linked data store will include pointers to both CRS and to Archie, it will allow queries that require data from both. An example would be the ability to have search results from CRS include Risk of Bias assessments from trials that have been included in a Cochrane Review.
- Allow for the flexible presentation of Cochrane content for a variety of publications and delivery methods. Because the metadata will point to specific components of Reviews, it will become possible to display selected portions of a group of related Reviews side by side.
- Improve searching and browsing of Reviews or CRS. The rich description of Cochrane data in the linked data store will allow faceted, tailored search on Cochrane's content directly by internal users such as our authors or editors. If our publishers incorporate the metadata into their own systems, it will also allow these improvements to be passed on to users of our products (clinicians, patients etc.).
- Allow third parties to align their metadata structures with those of the Collaboration through the use of common identifiers and compatible models – thus allowing an easier integration of our content with their own content.
- Allow linkages between specific pieces of Cochrane content and relevant material from the triple stores of other providers (examples: drug brand names or adverse effects, PubMed citations and related articles etc.).
- Improve efficiency in review production and authoring processes by providing access for authors to existing Cochrane data and facilitating its incorporation into their reviews.
- Enable methods research – especially research that involves comparisons across multiple Cochrane Reviews – as well as meta-epidemiological research.
- Facilitate the evaluation of Cochrane products and impact.

### Build Cochrane ontologies that defines key components of Cochrane content and their relationships with one another

---

This process involves two different types of activity. The more difficult is a process whereby the Collaboration and its contributors decide what the key components in our information architecture are, and how they fit together. This activity will require discussions throughout the Collaboration (which are likely to be very fruitful and useful in their own right). The models (ontologies) arising from the activity will be elaborated and modified gradually over time, a key benefit of using linked data technologies. Unlike most other approaches to data management, linked data operates on the "open-world assumption"<sup>2</sup> ([http://en.wikipedia.org/wiki/Open\\_world\\_assumption](http://en.wikipedia.org/wiki/Open_world_assumption)) and allows for expansion of the model(s) without breaking the system. The ontology developed for the demonstrator is a good starting place but will require some significant modifications as it was developed to serve only the user stories in the demonstrator and its logic reflects only the published review perspective and is not exhaustive in terms of systematic review methodology.

---

<sup>2</sup> Here, what we mean roughly is that all information and how we might want to work with a particular dataset or piece of content cannot be known ahead of time. So, the system has to allow for rapid integration of new models and ways of working and is thus more robust and future-proofed.

The second, technical activity of appropriately coding the results of this discussion (into URLs etc. as mentioned in the recommendations from Ontoba) can be done by our IT teams and is relatively straightforward, while the further development of the ontology itself will require a Knowledge Manager, along with consultants, to complete.

As the Cochrane ontology is being developed, it should be made public with clear documentation and versioning (for an example of documentation from a related project, see <http://rctbank.ucsf.edu/home/ocre>) and possibly uploaded and maintained in a tool like BioPortal, which “provides access to commonly used biomedical ontologies and to tools for working with them” (<http://bioportal.bioontology.org/>). This will allow others to comment on, enhance and collaborate with us in its development and to explore possibilities for linking with our data.

### Benefits to the Collaboration

---

This is an essential step in conceptualising and formulating a Cochrane linked data store. However, it has the potential to provide other significant benefits to the Collaboration, especially if done relatively quickly.

- Provide a clear reference point for communicating the Cochrane linked data work to both internal and external stakeholders.
- Allow us to develop structured mechanisms for others who might wish to interact with or license our data, or link to it from their own programmes and products.
- Provide a common framework, shared across the Collaboration for use in designing tools and processes to improve efficiency in review production and authoring.
- Provide leadership - A number of initiatives currently underway are attempting to define models (ontologies) for closely related areas, but no one to date has come forward with a semantic model for Systematic Reviews, for their key components such as risk of bias assessments or for the relationships between reviews, the trials they contain, and the multiple reports in the literature that may be associated with each trial. A Cochrane model that did so would stand a good chance of becoming a Web 3.0 standard approach.<sup>3</sup>

*"Just as companies who were quick to embrace the web, search marketing, and social media used that as a competitive advantage, a similar dynamic is likely to exist with the web of data.*

*Whose data will be the authoritative source in each market out there?*

*Whose data will be embedded in the most exciting third-party applications?*

*Whose industry-specific vocabularies will become de facto standards?*

*Whose open data will be most linked to within other data sets, generating a positive feedback loop around a data-level brand?"*

- From <http://chiefmartec.com/2010/07/data-as-a-new-marketing-channel/>

### Identify appropriate vocabularies or taxonomies from controlled and standardized sources to use in tagging our content with metadata

---

Once the pieces have been identified, and much of this work has already begun, it will be essential to develop agreement on classification systems (vocabularies, taxonomies) for those resources. These classifications will need to be agreed on and used across the Collaboration, and be structured in a way that allows them to be linked to other parts of the linked-data universe. Perhaps the most important of these external structures is the Unified Medical Language System (UMLS) developed by The US National Library of Medicine. It includes 135 “semantic types”

([http://www.nlm.nih.gov/research/umls/META3\\_current\\_semantic\\_types.html](http://www.nlm.nih.gov/research/umls/META3_current_semantic_types.html)), which are high-level categories covering more than 1 million concepts from the more than 100 vocabularies and terminologies (from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1435839/>). Within the UMLS,

---

<sup>3</sup> Web 3.0 is the larger movement within which linked data exists.

there are many taxonomies to choose from such as SNOMED CT, RxNorm, MeSH and DrugBank but no single taxonomy will fit all of our needs. In the demonstrator, for example, SNOMED and DrugBank worked reasonably well for classifying drugs that treat asthma. Another example is the BIBO ontology (<http://bibliontology.com/>) which was used in the demonstrator for citations of reviews and trials. In other cases, no existing classification scheme will exactly fit Cochrane's needs, and we may need to use a combination of existing schemes. Much of the "I" and "C" portions of PICO are likely to fall into this category. While there are several competing taxonomies for drug interventions, it is not clear that any of them exactly fit our needs. For example, most do not include "long-acting beta-agonists" (an intervention in over 30 existing Cochrane Reviews) as a category. In addition, there are no existing taxonomies that cover all of the non-pharmacological interventions addressed in Cochrane Reviews. In other cases there may be an opportunity for Cochrane leadership in the development of taxonomies to be used by others. For example, we have been unable to find a classification of medical outcomes and would see this as an opportunity to work with the COMET initiative (<http://www.comet-initiative.org>) to produce something of value to the entire health care linked data community.

Identification of appropriate vocabularies and taxonomies will require the expertise of TSCs and content experts in CRGs and Fields, and will build on the work already done by Cochrane groups in developing topic lists, and by the CEU in the development of the "browse" hierarchy currently used on *The Cochrane Library* and on [summaries.cochrane.org](http://summaries.cochrane.org). Some collaborative effort will be needed to shape these multiple, independently developed lists into a single coherent structure and an approach agreed in tagging the archive of all our content and in tagging new content going forward. There are several potential options here including having CRGs and/or authors do the tagging to having the tagging done centrally by the new Knowledge Manager post in coordination with staff from the CEU and informed by CRGs and others in the Collaboration.

### [Adopt a Service Oriented Architecture to improve inter-system communication](#)

This recommendation calls for the Collaboration to move from our current IT approach which is centred around individual IT systems, programmes or applications to one that is centred around the various "services" that we provide or may provide in future. To use a transportation analogy, our current approach is like building a railway system and keeping track of all of the train schedules, while the service oriented approach involves designing a highway system and letting people drive their own cars. The recommended design for this system is represented in the diagram on page 19 of the Ontoba report, and in the [figure 1](#) in the appendix of this document.

When our network involved only the IMS and Wiley, the railway approach worked well, but as the number of nodes in our network expands, a service oriented architecture (SOA) has a number of advantages. Proponents have suggested that a SOA can help organisations respond more quickly and more cost-effectively to changing conditions, helps promote reuse of resources and services in new contexts, can simplify interconnection to, and usage of, existing IT assets, and can improve the return on IT investments after the initial work of shifting to this approach is accomplished.

The thrust of this proposal is that the Collaboration's IT teams would develop "services" that could serve and accept data in an implementation-independent and non-proprietary format. This would allow interchange of data both across existing software systems within the Collaboration and also with the systems of others - but do so in a controlled and structured fashion.

Accepting this recommendation would not require a move to a new strategy: Some of this structure already exists within Archie and other Collaboration products, but it needs to be made more uniform and expanded. By taking a low risk and evolutive approach, the Collaboration can enhance our existing technical architecture with an expanded set of interfaces that will efficiently support key use

case requirements. These new services can be built in development iterations (cycles) based on the priorities we assign to individual user stories.

While the strategy is not completely new, its implementation will require significant work by the Collaboration's IT teams, who will need to maintain the functionality of our current systems while planning for and moving toward the new structure.

### Benefits to the Collaboration

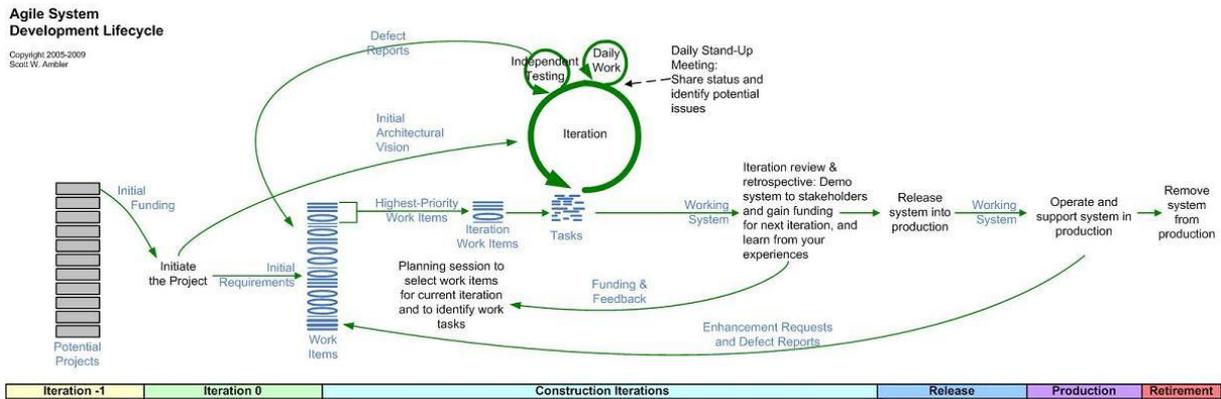
- Improve the interoperability between Cochrane's existing software and systems and allow for faster, more flexible, development and enhancement of our content.
- Improve the responsiveness of Cochrane software development activities to priorities identified within the Collaboration.
- Make it easier to make use of review production or management software developed by others for the authoring, editing and updating of Cochrane Reviews.
- Allow for the delivery of richer content services and more nimble content to enhance our current Reviews and derivative products and help us more-easily create new ones from our rich datasets.
- Reduce "vendor lock-in" with any one specific third party component. If all communication with the component is bound to a service API, then it is less costly and considerably easier to switch the component out for an alternative as only the service API needs to be changed to talk to the new component.

### Employ 'user stories' to identify and prioritise potential software development activities

In this approach, user stories provide a mechanism for an on-going dialog between user needs and desires and software development, with frequent (usually weekly) re-examination of the highest priority user stories and measurement of software development against the realisation of those stories. Thus, the anticipated process would be:

- ➔ High level development of Collaboration strategy - leading to:
- ➔ Identification of priority areas for software development - leading to:
- ➔ Specific projects focused around "services" to be developed or provided - leading to:
- ➔ An agile process that makes extensive use of user stories to develop services in a way that meets the highest priority user needs and wishes.

The following diagram (from <http://www.agilemodeling.com/artifacts/userStory.htm>) illustrates the process. The high level strategy and prioritisation processes determine which projects from the stack on the left are chosen for development. For each chosen project the relevant user stories ("work items") are used to iteratively drive software development. To give an example, if 'Create better and more widely used Summary of Findings Tables (SoFT), were identified as a project, the user stories might address disparate areas that included discoverability, display on *The Cochrane Library*, making it easier for authors to produce and update SoFTs, and development of specific SoFT derivative products. [Appendix 1 in our current list of user stories](#) has additional examples of multiple user stories around the theme of Cochrane Register of Studies (CRS) integration and Asthma review discoverability.



In addition to their role in iterative software development, user stories can also have a place in the higher-level prioritisation process. A user story that is suggested repeatedly, fits with other stories into a popular and coherent theme, or is compelling for other reasons, may lead to prioritisation of a particular project.

In order to implement this recommendation, the Collaboration will need to establish a process for eliciting user needs from throughout and beyond the Collaboration and shaping these into user stories that can appropriately inform the development process. We will also need to develop a framework for prioritising these work items in accordance with the Collaboration's technology, content and product strategies. These processes would build on the work of existing groups such as ADAC and RAC, but unlike those groups, would not restrict their focus to the capabilities of a single piece of software. Instead, the aim would be to be able to respond to user stories which (as in the examples above) have implications for and would draw data from several different Cochrane data stores.

### Benefits to the Collaboration

- Allows a flexible and nimble approach to software development that is responsive to user needs.
- Allows exploration and elaboration of related perceived needs across different groups in the Collaboration.
- Promotes a linkage between the Collaboration's vision and strategic goals and the perceived needs of the various Cochrane communities and stakeholders.

## 3. How will we benefit?

Linked data is a technology strategy that cuts across all elements of the work of Cochrane. It is as much about the way in which we create and store our data as it is about ways of using the data in new, creative ways. With this in mind we offer some examples of how various user groups in Cochrane will benefit in different ways.

### 3.1 Cochrane Collaboration

#### Systematic Review Ontology

We will be able to create, control and maintain a formal representation of systematic review concepts and the relationship between these concepts in a model known as an 'ontology'. This is important for a number of reasons as follows:

- It improves opportunities for Cochrane's leadership in systematic review methods to be maintained as the importance of linked data grows. Others are always free to create their

own ontologies, but the combination of Cochrane’s reputation and ‘first mover advantage’ means that others are more likely to use the Cochrane ontologies rather than creating their own. The translation of Cochrane’s methodological expertise into our ontologies benefits the wider systematic review community, and ultimately end users, because this will ensure that we remain leaders in this area.

- It ensures that Cochrane is able to shape the concept representations as per Cochrane’s understanding of systematic review methodological standards. This will help mitigate the risk of Cochrane coming under pressure to comply with alternative, less rigorous ontologies relevant to systematic reviews.

### Better delivery of content and interaction with third parties

---

- Flexible delivery of content makes derivative products easier to produce and manage.
- Being at the forefront of technology will position us well to take full advantage of future technological development.
- Interaction with third party software will allow for novel ways to increase exposure to our content, (e.g. PubMed could show a tag when a trial has been included in a Cochrane review, or it could show Cochrane ROB graphs alongside trial reports). This also opens opportunities to more explicitly work with others to disseminate our content, e.g. a charitable organisation could build a website that uses Cochrane data to provide evidence in their area of healthcare.

### Sustainability

---

- Using linked data can present opportunities to exploit novel income streams through selling access to our data. One real example came up recently through a request from a startup company in Hungary who are developing a diabetes management system and wanted to license access to our API (which we don’t currently have) for efficacy data from Cochrane to present alongside other information on their website. We were unable to help, but with linked data this would be a simple request, and such usage could be charged for.
- In addition to financial sustainability, with more efficient technical systems and better structure and enrichment of our content, we’ll be better placed to make use of our own content and data internally for our review production and evidence synthesis processes.
- The biomedical domain has whole-heartedly embraced linked data as the standard for data interoperability and for describing sub-areas of biomedical research via ontologies. Cochrane will fall quickly behind if we don’t adapt to this trend by being *the* dominant player in terms of evidence-based health care and systematic review methodology and production.

## 3.2 Entity staff

---

**CRGs:** better intelligence regarding our review portfolio (ability to query the system in any way to find out about review coverage, etc.), help with prioritisation, assist in creation of topic-based portals.

**Fields/Centres:** easier ways to find relevant content, e.g. Child Health could easily query the data for reviews that have relevance to children, including nuanced searches such as reviews including children where childhood studies were found; a Centre could easily find reviews that feature content specific to their region.

## 3.3 Authors

---

Linked data will provide multiple benefits for authors and improve the production of Cochrane reviews, both in terms of improving the efficiency of review production as well as the review quality. For example, linked data will enable easier identification of where studies have been included elsewhere, providing the previous RoB assessments made, highlighting where other authors have obtained translations or further information.

Furthermore, the authoring experience often requires substantial process recall. Capturing this process information enables authors to reproduce this quickly, meeting requirements of transparent reporting of the review processes, without it becoming a burden.

There would be significant benefits for overviews as linked data could facilitate with identification, retrieval and re-use of study data from more than one review for the purpose of generating a dataset relevant for overviews and network meta-analysis.

### **3.4 Users of Cochrane content**

---

Linked data will allow us to disseminate our content far more effectively, in various different ways which will facilitate improved access to our content. It will also allow us to repurpose content, work with third parties and deliver specific pieces of information for use outside of Cochrane. Better and more intelligent searching of Cochrane content will allow users to restrict searches by diverse parameters, from age groups, to outcomes, to a specific group of drugs etc., and being able to search by trade names. Flexible delivery means that we can present our content in alternative ways more easily, e.g. apps, customisable presentation of reviews, etc., so that individual users can access our content in a way that suits them.

## **4. Challenges of the project**

---

Implementing this new technology framework for the Collaboration will not be without challenges and there will need to be substantial thinking and resources applied to “change management”. These include, but are not limited to:

### **4.1 Organisational and cultural challenges**

---

The initial challenge will be in communicating the benefits of the work effectively to Cochrane users, both internal and external. As the change to using linked data can be conducted in an incremental fashion, it will be easy to see some smaller changes fairly quickly, but less easy to see the possibility of ‘finding what you don’t know you want to know’! However, the adaption of the focus on user stories should also see some fairly quick benefits to users.

### **4.2 Technology teams and our Publisher**

---

There will need to be some investment in supporting the technology teams to improve their ways of working together and to work on unified approaches for the new developments. Changes to this effect are in motion as there are many obvious benefits for the technology teams working more closely together to support more efficient working, generally.

The changes we make as an organisation will impact our publishers, and our publishers will also need to be on board and ready to adapt to our new ways of working. This means that it will be essential that we work closely with our publishers as we make these changes in order that they can present the information well to our users.

### **4.3 New ways of tagging and displaying our content and data might bring concerns**

---

The change will also have an impact on other Collaboration employees. The tagging of archive and new records may be undertaken by CRGs and/or authors or maybe performed centrally by the Knowledge Manager and may be semi-automated. Drawing on the expertise of TSCs who are already familiar with controlled vocabularies in their fields many CRGs may ask TSCs to be involved in this process. In those CRGs where the Trials Search Coordinator (TSC) undertakes this additional role it is likely to have an impact by including new ways of tagging and displaying content, and adding metadata via controlled vocabularies in the biomedical domain. Ultimately this should support the

work TSCs currently do, and increase inter-entity communication and improve abilities to retrieve data stored within other entities. The work will require additional support to the TSC role and changes will be introduced in an incremental fashion. Those CRGs and TSCs where the impact on workload would be significant will need to be reassured that provision would be provided centrally by (or via) the Knowledge Manage and CEU. Providing flexibility of approach where, for example, some keywords may be added to a record when it first enters the CRS and then later addition of keywords by review authors and others on retrieval of the full text of the paper may be the best approach.

#### **4.4 Methodological issues need to be dealt with**

---

Any changes linked data bring are likely to have an impact on Methodology for the organisation, both in terms of how to conduct reviews, and how to use information that will be retrievable by having access to linked data. For example being able to use and retrieve methodological data that has been created by another user will have to come with guidance from experts in Methodology in Cochrane. And, creating new interfaces and formats for our content needs to be done carefully so that the context and provenance of the information is retained and no methodological errors or misleading views of our content are introduced.

#### **4.5 Managing expectations**

---

It will be important to manage expectations within the organisation. Linked data will not be a 'fix all', but it will improve efficiency for retrieval of data, assist us in responding to user stories within the organisation, as well as giving us support to more easily find solutions to the queries that have 'yet to be asked'. Additional staff will inevitably be required to support the work, but aside from two initial posts, any additional staff will be sought in an incremental fashion, when required through the course of the project. It will need to be made clear that the introduction of linked data will not mean that our normal work stops, for example the development of Cochrane.org and RevMan will both continue mostly as normal, with the additional benefits from introducing linked data following incrementally.

#### **4.6 Evaluation - "How will we know we've been successful?"**

---

Due to the iterative and incremental implementation process, there should be small wins earlier on with the introduction and exploration to develop answers for user stories, with more significant changes rolling out in the longer term and dependent on our publisher's capabilities. Due to the move to a more agile or responsive way of working required for the move to linked data, development work and time should be reduced thus facilitating more efficient delivery to the end-user. As the workload will continue to increase as the organisation continues to expand, the resultant effects on staff may be less significant than with the current structure, thus supporting increased sustainability long-term with increased value to the user. So we should be setting ourselves up in the future as a much more responsible and efficient organisation with more responsive and effective technology to support our work. This is how we will know this project is successful.

## **5. What's the larger context?**

---

Technology has always been a key enabler of systematic reviews and many technologies are embedded in current Cochrane Review workflow, including online bibliometric databases, reference management software, Archie and RevMan, as well as generic tools such as word processing, spread sheets and email. Although these internal and external software systems continue to evolve, the pace of development is now accelerating. This has many potential benefits for Cochrane, but also significant risks.

An example of the potential benefits of the acceleration in technological innovation for Cochrane is the development of tools that support the portion of review workflow not addressed by Archie and RevMan, such as Distiller SR, EPPI-Reviewer, ReGroup and EROS. These tools improve the user experience and efficiency of review production, but within the context of linked data, raise the question of how Cochrane best interacts with these products to ensure the evolving needs of a linked data approach are able to be met. At present, these tools seek to support user needs, which include the ability to import data extracted from study reports into RevMan. However, these tools do not currently interact smoothly with RevMan and in the future it is likely that additional integration and data flows will be important. The Collaboration will need to develop ways of managing these data requirements as needs develop.

More broadly, the Collaboration operates in an increasingly crowded space occupied by other evidence-based practice organisations. A linked data approach holds promise of richer ways of interacting with these organisations given that the very essence of linked data is the interoperability of data across organisations and systems. At the same time, it is likely that the development of systems that enhance the ability to discover, organise, summarise and deliver health data will proliferate. For example, the Systematic Review Data Repository (SRDR) is being utilised by the Agency for Healthcare Research and Quality Evidence-based Practice Centers (AHRQ EPCs) to collect data in a single repository. Over time, initiatives such as these will have significant impact and the Collaboration's leadership will continue to be dependent on innovations that keep Cochrane at the forefront of systematic review methods and practice.

Perhaps most importantly, it is increasingly clear that very large sets of data – and how these data are stored, used and governed – will increasingly shape knowledge intensive sectors such as health. Linked data is an approach that is growing rapidly and nowhere more so than in the life sciences. The benefits of these changes will be significant, but just as other sectors such as media, communication, retail and finance are being disrupted by technology, so increasingly will the health sector. In these other sectors, many established organisations have embraced the rapid change brought on by technology, whilst others have found their business models rapidly superseded. In the field of evidence for health, Cochrane is ideally placed to lead, disrupt and benefit from technological innovations and simultaneously at risk of being left behind as new models are developed by others.

## 6. How will we proceed?

---

### 6.1 Gradual iterative approach

---

One advantage of the linked data approach is its flexibility and extendibility. This means that the entire linked data system does not need to be designed and implemented all at once. A commitment to a linked data approach would still allow the Collaboration to make use of our existing software and data stores, but would call for us to gradually extend and enhance them with linked data. We would propose an implementation plan with 3 phases:

- **Foundation** - This important first step will put into place the necessary people and systems to allow The Collaboration to address user stories with a linked data approach.
- **Exploration** - In this phase, a small, carefully selected set of user stories will be addressed with the aim of exploring the applicability of linked data to specific situations while building and expanding the Cochrane ontology, vocabularies and linked data processes.
- **Production** - Eventually, linked data approaches can be used to address mission-critical components of review production, publication and access by users.

The complexity of user stories and the pace at which they are addressed in phases 2 and 3 can be varied depending on existing resources and our growing experience with linked data.

## Foundation Stage - Establishing basic linked data capabilities

---

This section outlines some basic requirements which must be in place in order to begin addressing Cochrane user stories using a linked data approach.

### Hire or contract with key personnel

---

#### Chief Technical Architect (CTA)

---

As this person will lead and direct the change to our current technical architecture and to evaluate our interaction with our current technology systems, it is critical to fill this role as soon as possible with someone who understands the process of moving a complex organisation to a linked data approach. Having a CTA on board will help us to begin with the most important steps and to do things in the right way, right from the start. It may be possible to hire someone with the necessary expertise and experience on a time-limited contract to get us off on the right foot and then train and pass on the on-going maintenance activities to someone less experienced once the initial design and structure decisions have been made.

#### Knowledge Manager

---

This person would be responsible for working with the various Cochrane groups to help them in moving toward common vocabularies – starting with an assessment of what we already have and assessing strengths, weaknesses, and areas of overlap, then beginning to explore linkages with vocabularies and taxonomies developed by other organisations and used in other systems. In addition, the Knowledge Manager would manage these metadata sets and oversee the annotation of Cochrane content and maintain the suite of Cochrane ontologies (semantic models of our content and data) and adapt them as needed to fulfil user stories and larger use cases.

#### Consultants

---

We are likely to benefit from consultants with expertise in several different areas.

- Moving organisations to linked data and identifying big picture issues in moving forward. This will overlap with the CTA role but can provide an additional perspective, and be a resource to the CTA. Ontoba has done this for us so far.
- Use of linked data to medical literature related applications – to provide or steer us to relevant tools for building our ontology, annotating our Reviews etc. Alexander Garcia Castro (<http://biotea.idiginfo.org/>) has done more on this than anyone else we have identified so far.
- Existing ontologies and how they might map to our needs. Could supplement the work of our Knowledge Manager and save that person a lot of time and missteps. Richard Boyce has provided valuable information on drugs ontologies.
- Linked data approaches to open publication, data repositories etc. Ida Sim would be one possibility.

### Buy or license the needed hardware & software

---

Some hardware upgrades as well as new software and upgrades to existing software will be needed. We plan to cost these out during the “Foundation” phase and to ramp up, as needed, with purchasing and installing these components. For example, as the “Foundation” and “Exploration” phases won’t require production-ready servers and systems in our technical architecture, we can keep costs as low as possible during these phases and then add necessary components and perform the needed upgrades as we “go live” with using these new technologies and systems.

#### Cost considerations

---

We don’t expect these costs to be exorbitant and the triple store software we purchased last year for the demonstrator phase of the project, OWLiM, was offered to us by its maker, Ontotext

(ontotext.com) at a half price discount as we are a non-profit. They have indicated that any upgrades will enjoy this same discount. Upgrades to our server infrastructure will need to be assessed when we determine what capabilities will “go live” first in the rollout.

---

### [Begin work on core Cochrane linked data structures](#)

---

#### The Cochrane Ontologies

---

*“To approach semantic technology at the strategic level, then, start by getting on-board with an enterprise ontology, “It creates context and a road map that lets you make sure each initiative you do going forward is additive to move you to the future rather than be another neo-legacy system that gets in the way later,” McComb advises. “One of the great things about having an enterprise ontology is that you can use it in the short-term as just a more rational way to communicate between existing systems.”*

*Expect to spend four to six months creating a first version, and then set in place a governance structure to recognize and account for the fact that that version will evolve as it gets used. Members of that governance body can include representatives from each business unit and have some in-house talent to take change requests and introduce them to the enterprise ontology. This approach “also presents a great opportunity for people to rationalize their taxonomy initiatives,” says McComb. “Many people have them and they are disconnected. With an enterprise ontology it tells you where each taxonomy fits into the larger model and it tends to lead you to smaller and more reusable taxonomies.”*

- From [http://semanticweb.com/for-the-enterprise-it-set-steps-to-success-with-semantic-tech\\_b36456](http://semanticweb.com/for-the-enterprise-it-set-steps-to-success-with-semantic-tech_b36456)

---

#### Keep the “demonstrator” in operation and use it as a “sandbox”

---

The [demonstrator.cochrane.org](http://demonstrator.cochrane.org) site has served this purpose up until now. By leveraging the linked data modules and capabilities within our content management system (CMS), Drupal, we can quickly and easily test new presentation and “views” of our content in an iterative, low-risk development environment. This will help us make decisions about priorities for both our published products, and thus will inform discussions with Wiley and other potential “downstream” consumers and publishers of our data, but also can serve as source for “data intelligence” and development of tools to assist in the authoring process.

---

### [Strengthen existing Cochrane teams and build partnerships with collaborators](#)

---

#### Consider paying for time from Cochrane contributors

---

The move to linked data will require a fair bit of change from CRGs and other Cochrane Groups. We should consider a mechanism for either doing the work for them with linked data funding or providing them with funding to accommodate the extra workload – especially during the early phases. As the project progresses, we should start decreasing their workload, so this funding will probably no longer be necessary.

---

#### Expand the capabilities of the Collaboration’s software teams

---

Training or new hires to bring needed linked data expertise. Provide sufficient capacity to allow development of new software, maintenance of existing software, and exploration of possible new approaches to specific user stories using linked data.

---

#### Identify key potential partners/collaborators and begin to explore possibilities for collaboration

---

CRD, Ontotext, Epistemonikos, Norwegian Knowledge Centre (Andy Oxman, Linn Brandt, Sarah Rosenbaum, DECIDE, MAGIC, etc.), Ilkka Kunnamo/Duodecim, <http://biotea.idiginfo.org/>, EPPI-Centre (James Thomas), Brown University (Byron Wallace, Tom Trikalinos)... The list keeps growing.

### **Exploration Phase - Using linked data to address selected user stories**

---

Once this basic linked data infrastructure is in place, it will be possible to begin using it to address selected user stories. These initial user stories will serve a dual purpose. In addition to meeting the needs of the users who proposed them, they will be used to guide the on-going development of Cochrane linked data structures. It is anticipated that the initial "users" would be internal. Examples might be editors wishing to have quick access to selected portions of a review to check for fit with MECIR standards (see the [Cochrane Quality User Story](#)), Fields wishing to explore the fit between selected reviews and the needs and interests of their constituents (see the [Child Health User Story](#)), methodologists wishing to examine specific characteristics of a set of reviews, or TSCs wishing to make the contents of CRS easily accessible and useful for review authors (see the [PubMed ID User Story](#)). In each case, the users would be asked to work with the Chief Technical Architect and the Knowledge Manager to determine the fit between their needs and the existing Cochrane linked data structures and to explore ways that the Cochrane ontology could be expanded to meet the needs of their user story and of similar stories likely to be presented by other users in future. The very first efforts in this phase would be fairly simple and narrowly focused around selected components of Cochrane data, but the linked data structures built for each user story would be re-usable for subsequent stories, thus allowing gradually increasing complexity and sophistication of the user stories.

### **Production phase**

---

While the first two phases could in theory be accomplished by the Collaboration working alone, this phase will require joint activities with our publisher. (In fact, it will be important to at least make our publisher aware of our activities in the first two phases in order to be certain that the Cochrane linked data structures being developed are compatible with directions being taken by the publisher in their own software development plans.) User stories chosen for this phase will involve modifications to the ways Cochrane data are exposed to users (either in reviews or in derivative products), and the programming involved will require collaborative efforts between the software teams of the publisher and those of the Collaboration. While these user stories will necessarily be more complex than those in Phase 2, it will still be possible to begin with some of the simpler ones and gradually progress through more challenging cases using the same iterative approach outlined for phase 2. Some early user stories might involve the ability to search *The Cochrane Library* by drug brand name (already implemented on [summaries.cochrane.org](http://summaries.cochrane.org), although not fully using a linked data approach), or to use the Cochrane ontology to allow users to better focus their search (e.g. "show me reviews assessing options for patients with type 2 diabetes and cardiovascular comorbidity who are already taking metformin").

### **Importance of evaluation at every stage**

---

Throughout all three phases, evaluation will be critical. The agile, iterative approach to development, coupled with the user-centred design approach, should ensure that we are developing our systems to meet users' needs, both in review production but also with regards to end-users of our products and services. However, it will be of utmost importance that methodological expertise is present in evaluating this development to ensure that our content remains high-quality and reliable.

## **8. Summary**

---

*"Information and Communication Technology (ICT) is at the heart of The Cochrane Collaboration*

and is an essential support to our primary purpose of producing high-quality systematic reviews. It is a crucial element in the way that Cochrane Reviews are generated, it facilitates communication within the Collaboration, it is the means by which people find out about the Collaboration and its work, and it is a fundamental requirement for users of The Cochrane Library. To a very large extent, the success of The Cochrane Collaboration has been based on its investment in ICT.”

- From the “Introduction” to “A Collaboration-wide Information Services Strategy Framework” (<http://www.cochrane.org/community/organisation-administration/information-services>)

“...relational databases don’t combine easily. Linked Data is a standard international mechanism for performing data inter-operability across systems, so the potential for the enterprise is to use Linked Data to save a lot of time, effort and money and let data inter-operate across silos cheaper and faster. That’s a big deal especially in times of decreasing capital budgets, CIOs ignore this at their peril.”

- David Wood, CTO, 3 Round Stones

Linked data will be core to the #CochraneTech strategy to 2020, and beyond. Moving Cochrane’s technical infrastructure, teams and systems to the approach outlined in this paper will ensure that we are “future-proof”, more sustainable in the production and delivery of our content and more “nimble” (<http://nimble.razorfish.com/publication/?m=11968&l=1>). The underlying logic and motivation behind this project is in line with both Cochrane’s strategic goals and its core principles. This work will align with and build on the vision laid out in the *Information Services Strategy Framework and Structure* (<http://www.cochrane.org/community/organisation-administration/information-services>) drafted in 2011.

## Our principles

---

This new technical strategy built within a linked data framework maps to and supports all of our core principles.

### ➔ Collaboration

*by internally and externally fostering good communications, open decision-making and teamwork*

Linked data will join Cochrane to the “web of data” and allow us to collaborate more effectively within the organisation, but also, and critically, it will allow us to collaborate with external organisations and projects in evidence-based health care via open standards in data interoperability. Cochrane will position itself to lead and be the standard in this technological space.

### ➔ Building on the enthusiasm of individuals

*by involving and supporting people of different skills and backgrounds*

The new technical architecture built around linked data builds on the progress of the last twenty years and, while it will be transformative, will expand on and not replace what we have, in terms of systems, data and methodology to further our mission.

### ➔ Avoiding duplication

*by good management and co-ordination to maximise economy of effort*

The linked data approach will mean we’ll have a much better handle on where content and data already exist and, via tagging and annotation, how to retrieve and see what content is related to aid in re-use and avoiding duplication of effort.

### ➔ Minimising bias

*through a variety of approaches such as scientific rigour, ensuring broad participation, and avoiding conflicts of interest*

Better data management will enable process recall which will lead to higher quality reporting which can minimise bias. Also, having better intelligence about our data, which studies have already been assessed and included in our Reviews and data-sharing can also help, as well as enhancing transparency in methods and reporting.

➔ **Keeping up to date**

*by a commitment to ensure that Cochrane Reviews are maintained through identification and incorporation of new evidence*

This new technical architecture will allow for our current systems to talk to each other more effectively but also for us to talk to external systems better. Linked data combined with technologies such as text mining and machine learning will mean we can identify new studies for incorporation into our reviews more quickly thus keeping the evidence more up to date.

➔ **Striving for relevance**

*by promoting the assessment of healthcare interventions using outcomes that matter to people making choices in health care*

Linked data will allow Cochrane to join the “web of data” and to promote our content more in both the online publishing and content delivery space, but also in the emerging “big data” and “open data” movements.

➔ **Promoting access**

*by wide dissemination of the outputs of the Collaboration, taking advantage of strategic alliances, and by promoting appropriate prices, content and media to meet the needs of users worldwide*

This is probably the strongest link between this proposal and one of our core principles. By using the linked data approach outlined in this document, we will be able to both improve and promote better access to the outputs of the Collaboration and open up new dissemination channels and revenue streams.

➔ **Ensuring quality**

*by being open and responsive to criticism, applying advances in methodology, and developing systems for quality improvement*

In short, better intelligence about our data and honing and encoding our methodology via ontologies and linked data structures, combined with an agile, iterative approach to development that is user-centred, will allow us to be responsive, nimble and to develop systems that are better quality-controlled and easier to manage.

➔ **Continuity**

*by ensuring that responsibility for reviews, editorial processes and key functions is maintained and renewed*

This new architecture will enable the creation of better processes and workflows for editorial review and maintenance of our content.

➔ **Enabling wide participation**

*in the work of the Collaboration by reducing barriers to contributing and by encouraging diversity*

The new technical architecture built around linked data will enable the better integration of software, websites and tools for systematic review production and dissemination thus enabling more people in more contexts worldwide to participate and contribute to the work of The Cochrane Collaboration.

## **Our strategic goals**

---

This new technical strategy built within a linked data framework will support all five of our strategic goals.

- ➔ Goal 1: To ensure high quality Cochrane systematic reviews are available across a broad range of healthcare topics.

This project seeks to improve review production and evidence synthesis via better and smarter use of technology.

- ➔ Goal 2: To promote access to Cochrane reviews and the other products of The Cochrane Collaboration.

This project seeks to improve dissemination, discoverability and delivery of Cochrane content via better and smarter use of technology.

- ➔ Goal 3: External focus.

This project introduces a system-wide user-centred design approach that looks externally to gather user stories and to meet the needs of our various end-users. In addition, the Collaboration will be joining many others in the biomedical research domain who have already adopted linked data and are seeking information that provides quality assessment and validation such as Cochrane evidence.

- ➔ Goal 4: To ensure an efficient, transparent organisational structure and management system for The Cochrane Collaboration.

This project has the potential to aid in better processes and workflows within the organisation to enable more efficient working in producing Cochrane Reviews via technology. In addition, and critically, this project will draw the two technology teams, Web and IMS, closer and create an architecture that will enable more efficient and effective working as well as promoting greater transparency and openness within a clearer structure.

- ➔ Goal 5: To achieve sustainability of The Cochrane Collaboration.

This project has the potential to open Cochrane up to new revenue streams and business models and to enhance our current products thus contributing to the financial sustainability but also in strengthening the Cochrane brand and ensuring that we remain the trusted source and market leader in evidence-based healthcare knowledge.

## **Building on the Information Services Strategy Framework and Structure**

---

From the “Core Principles” section of “A Collaboration-wide Information Services Strategy Framework” (<http://www.cochrane.org/community/organisation-administration/information-services>):

*The development and operationalisation of our Information Services strategy must:*

- 1. Adhere to the Collaboration's ten key principles, since they are based on inclusiveness, communication, transparency, efficient working, and methodological rigour;*
- 2. Be accountable to our Steering Group (board of trustees), facilitated by the newly established Information Services management structure;*
- 3. Demonstrate:*
  - Value for money*
  - An awareness of the potential of ICT to support and promote our work*
  - The prioritisation of high-quality, innovative, flexible and integrated technical solutions in order to provide a robust Information Services infrastructure, whether they are developed 'in-house' or acquired from an external provider*
  - Good governance*
  - Accountable and business-like decision-making*
  - Flexible approaches to problem solving*
  - Inclusiveness: in particular to create clear pathways between strategy development and substantive results for users, both Collaboration contributors and product end-users*
  - Accessibility: to ensure technology is developed and implemented to enable global use*

From the “Key users and their needs, current and future” section of “A Collaboration-wide Information Services Strategy Framework” (<http://www.cochrane.org/community/organisation-administration/information-services>):

- People who prepare Cochrane Reviews*
- People within the Collaboration who support those preparing Cochrane Reviews*
- People within the Collaboration who support the running of the organisation*
- People who wish to learn about the Collaboration and its work*
- People who wish to make evidence-based decisions in health care and policy*

This new technical architecture built around linked data adheres to our ten principles and will demonstrate value for money via better intelligence, use and re-use of our data and enabling software and systems, both internal and external to Cochrane (our publisher and potential external consumers of our data and content), to operate more efficiently. The linked data project grew out of an awareness of the potential of ICT and we have been careful to maintain good governance throughout the prototype and demonstrator phases of the project thus far. The process has been driven by business cases and user needs and not by technology and has operated using flexible approaches to solving these business and user needs. Inclusiveness has been core to the project thus far and will continue to be going forward by taking a user-centred design approach and interacting with users, both internal and external, of our software, systems and websites. In addition, linked data enables much greater accessibility and interoperability of our data and content and will ensure as wide as possible global use. Finally, this new approach will address the needs of and assist at some level all five of the key users identified in the Information Services Strategy Framework.

## Appendix

---

---

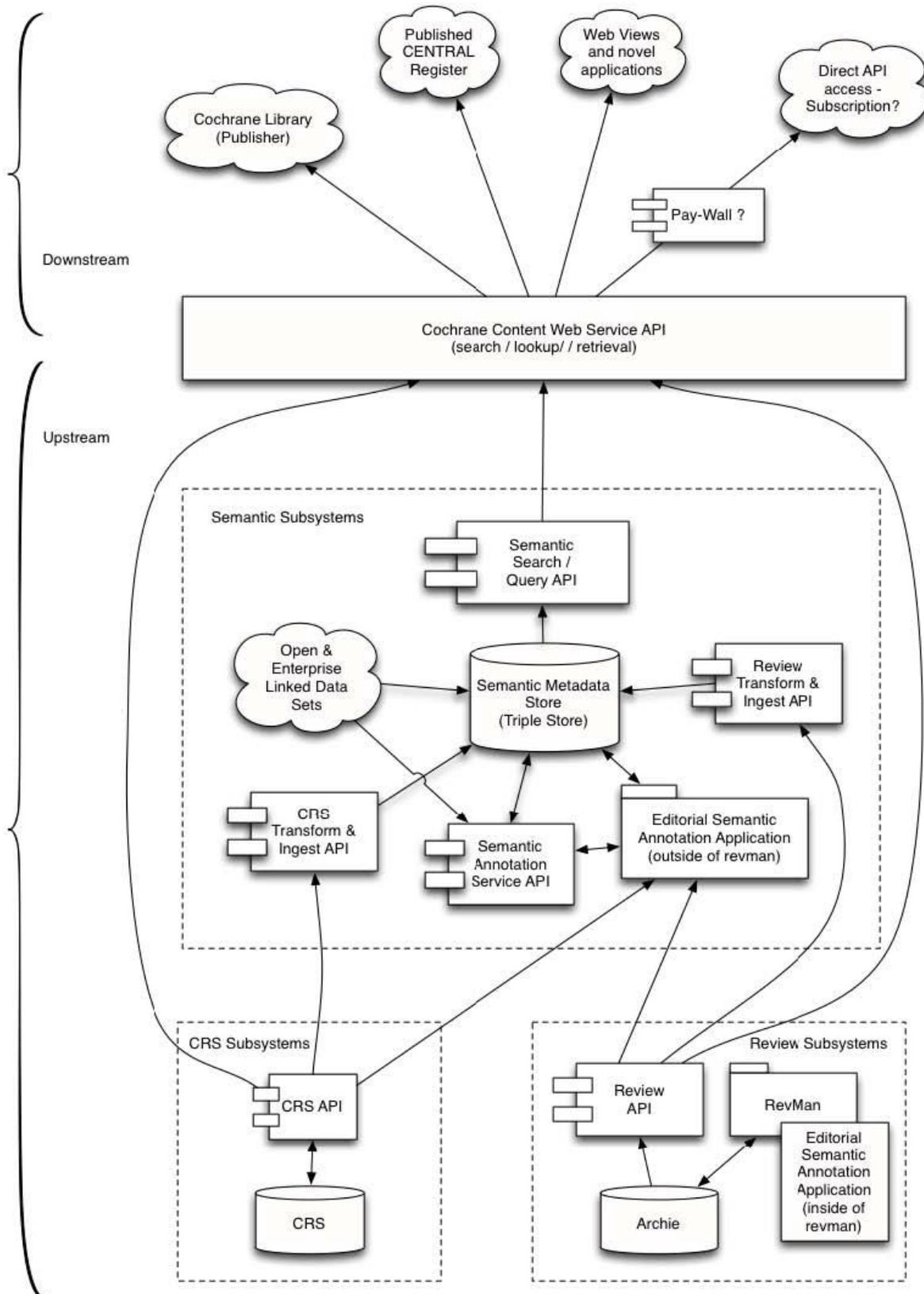
### Related documents – Attached as separate documents

---

**Linked Data Background paper: From December 2012 meeting in London**

**Ontoba report: *Linked Data at The Cochrane Collaboration: A technical strategy***

---



**Figure 1 - Linked Data Service-oriented Architecture: Diagram from the Ontoba Report**

