

Conceptos de Ingeniería de Software Basada en Evidencias

Sebastián Pizard, Fernando Acerenza, Vanessa Casella, Silvana Moreno
Rodrigo García, Joaquín Lezama y Diego Vallespir
Instituto de Computación, Universidad de la República, Montevideo, Uruguay
Contacto: spizard@fing.edu.uy

Resumen—Este documento presenta una guía básica a la Ingeniería de Software Basada en Evidencias. Se describen sus objetivos, su métodos más utilizados y además se incluyen referencias para profundizar en cada tema específico.

I. INTRODUCCIÓN

Este reporte¹ tiene como objetivo presentar los fundamentos de la Ingeniería de Software Basada en Evidencias (EBSE, por sus siglas en inglés), y en particular se incluye una breve guía para realizar revisiones sistemáticas en ingeniería de software. Este documento puede ser utilizado como guía inicial o referencia por estudiantes o investigadores que se encuentren realizando trabajos de EBSE con el Grupo de Ingeniería de Software (GrIS). Este documento se actualizó para tener en cuenta el libro de referencia de EBSE publicado en 2015 [1], en algunos temas se indican qué capítulos del libro se pueden consultar para profundizar más. En la sección II se presenta la definición de la EBSE así como sus motivaciones. En la sección III se describen las revisiones sistemáticas de literatura, en la sección IV los estudios de mapeo. En la sección V se introducen las revisiones terciarias. Más adelante, en la sección VI se presentan las revisiones rápidas (*rapid reviews*). Por último, la sección VII incluye una breve introducción a la traducción de conocimiento.

II. INGENIERÍA DE SOFTWARE BASADA EN EVIDENCIAS

La ingeniería de software basada en evidencias tiene como propósito mejorar la toma de decisiones relacionada al desarrollo y mantenimiento de software integrando la mejor evidencia actual de la investigación con experiencias prácticas y valores humanos [2].

En general, el conocimiento se deriva de la evidencia a partir de un proceso de interpretación [1]. Esto ocurre, por ejemplo, cuando un científico estudia registros médicos para demostrar que fumar tabaco causa cáncer de pulmón. En áreas con investigación empírica o experimental la evidencia se obtiene mediante observaciones y mediciones, cuyos resultados son reportados en una o más publicaciones científicas.

En el pasado relativamente cercano, para disponer del conocimiento más actual sobre un tema era común que un experto hiciera una reseña de las publicaciones que creía más relevantes y actuales. Estas revisiones tradicionales o reseñas

tienen sesgos relacionados a la experiencia del investigador, a su subjetividad y además no son del todo reproducibles.

Un caso muy famoso que muestra el sesgo del investigador es la revisión de Linus Pauling de 1970 sobre los beneficios de la vitamina C para combatir la gripe [1]. En esa revisión el autor tuvo en cuenta las publicaciones que apoyaban su teoría y descartó el resto sesgando fuertemente los resultados.

Al agregar los resultados de varios estudios, pueden surgir problemas, por ejemplo, al tratar de decidir cuándo hay que descartar evidencia débil. Para mejorar la agregación de evidencia surge para medicina en la década de 1970 la práctica basada en evidencias (EBP, por sus siglas en inglés). La EBP busca utilizar un enfoque objetivo, riguroso y planificado para seleccionar estudios relevantes y realizar una síntesis de los resultados de esos estudios [1]. La rigurosidad metodológica hace que los resultados sean más confiables ya que es posible estudiar el procedimiento llevado a cabo para su obtención así como también reproducirlo. En medicina la EBP ha sido fundamental para ayudar a controlar los factores de riesgo de infarto de miocardio y accidente cardiovascular, para transformar el VIH de una infección mortal a una crónica, para probar medicamentos para la hepatitis C y para mejorar tratamientos de algunos tipos de cáncer [3].

Las técnicas utilizadas en EBP son llamadas estudios secundarios, ya que realizan la agregación de evidencia a partir de estudios primarios (experimentos controlados, estudios de caso, encuestas, entre otros). El principal estudio secundario es la revisión sistemática de la literatura (SLR, por sus siglas en inglés). Las SLRs permiten recolectar y sintetizar evidencia de distintas fuentes. La característica clave que las distingue de las revisiones tradicionales narrativas (o clásicas) es su intento explícito de minimizar las posibilidades de llegar a conclusiones erradas, que puedan resultar del sesgo en los estudios primarios o en el proceso de revisión [4]. Para lograrlo se debe establecer un plan, llamado protocolo, con todas las actividades a realizar y criterios a utilizar previo a la ejecución de la SLR.

La introducción de la EBP en el área de la ingeniería de software comenzó en 2004 [2], llamándose EBSE, y tuvo una gran aceptación por parte de los investigadores. Se estima que se han publicado más de 200 estudios secundarios solamente en los primeros diez años [1]. Muchos de sus resultados se han utilizado como punto de partida para investigaciones más amplias. EBSE tiene tanta importancia en la investigación

¹Versión 2.0 - Julio 2019

hasta el punto que algunas revisiones han cambiado el sentido común dentro del área. Por ejemplo, Jorgesen et al. en 2004 no encontraron evidencia que apoye que los modelos de estimación son mejores que la estimación por expertos. Notaron que hay situaciones en las cuales los modelos no incluyen información importante del dominio de aplicación que si es tenida en cuenta por los expertos [5].

La ingeniería de software basada en evidencias alienta un fuerte énfasis en el rigor metodológico involucrando los siguientes cinco pasos [2].

1. Convertir un problema relevante o una necesidad de información en una pregunta que pueda ser respondida.
2. Buscar en la literatura la mejor evidencia para responder a esa pregunta.
3. Evaluar de forma crítica la validez, el impacto y la aplicación de la evidencia.
4. Integrar la evidencia evaluada con la experiencia práctica y los valores y circunstancias de los interesados.
5. Evaluar la efectividad y la eficiencia de este proceso para buscar maneras de mejorarlo.

Los primeros tres pasos son esencialmente el papel de la revisión sistemática de literatura (SLR), mientras que el cuarto es el de la traducción del conocimiento. El quinto es garantizar que los procedimientos de investigación estén sujetos a un análisis constante.

Existen otros métodos que pueden ser vistos como adaptaciones de la SLR, estos son: estudios de mapeo, revisiones terciarias y revisiones rápidas. Las siguientes secciones introducen a estos métodos, profundizando más en las SLRs e indicando para los otros métodos en qué se diferencian. Por último, se incluye una sección que introduce brevemente a la traducción de conocimiento.

Dos aspectos importantes a tener en cuenta al usar EBSE son:

- Características de la ingeniería de software — Los estudios primarios implican participación activa, la ingeniería de software carece de taxonomías fuertes, los estudios primarios carecen de poder estadístico, hay muy pocos estudios replicados, los estándares para reportar investigación son a menudo deficientes.
- Limitaciones de EBSE — Las personas realizan una revisión sistemática, los resultados dependen de los estudios primarios, no todos los temas se prestan bien a los estudios empíricos

Sobre este tema:

- Los capítulos 1 y 2 de [1] presentan una muy buena introducción al paradigma basado en evidencias ya a EBSE respectivamente.
- También es posible profundizar en características de la ingeniería de software y las limitaciones de EBSE en las secciones 2.5 y 2.6 de [1, cap. 2].

III. REVISIONES SISTEMÁTICAS DE LITERATURA

III-A. Motivación

La necesidad de agregar evidencia desde múltiples estudios empíricos tiene varias motivaciones. Seguramente al comenzar una investigación se quiera utilizar toda la investigación previa relacionada con el fin de no partir de cero. También es posible querer contemplar los resultados de varios estudios particulares en conjunto para contestar preguntas amplias que no podrían contestarse de forma individual. Este proceso de recolección y síntesis debe contar con cierto estándar científico y rigurosidad metodológica.

Las revisiones sistemáticas de literatura permiten recolectar y sintetizar evidencia de distintas fuentes. La característica clave que las distingue de las revisiones tradicionales narrativas o clásicas (aquellas sin un enfoque metodológico claro) es su intento explícito de minimizar la posibilidad de llegar a conclusiones erradas, que puedan resultar de sesgo en los estudios primarios o en el proceso de revisión [4].

Sobre este tema:

- Una breve descripción de las revisiones sistemáticas de literatura en ingeniería de software puede encontrarse en [6] o en [7, cap. 4], la referencia más importante actualmente es el libro de Kitchenham [1].
- En [8] se presenta una breve pero útil introducción a los distintos métodos empíricos utilizados en la investigación en ingeniería de software.

III-B. Definición y Etapas

Una revisión sistemática de literatura (o SLR por sus siglas en inglés) es un método para identificar, evaluar e interpretar todas las investigaciones pertinentes a una determinada pregunta de investigación, área temática o fenómeno de interés. Los estudios individuales que contribuyen a una revisión sistemática son llamados estudios primarios; mientras que una revisión sistemática es por tanto un estudio secundario [9].

Una SLR tiene ciertas etapas discretas contempladas en tres fases principales [9]:

1. Planificar la Revisión
 - Especificar las preguntas de investigación.
 - Desarrollar el protocolo de revisión.
 - Evaluar el protocolo de revisión (recomendada).
2. Realizar la Revisión
 - Identificar la investigación.
 - Seleccionar los estudios primarios.
 - Evaluar la calidad de los estudios.
 - Extraer datos y monitorear.
 - Sintetizar los datos.
3. Informar la Revisión
 - Especificar los mecanismos de difusión.
 - Formatear el informe principal.
 - Evaluar el informe (recomendada).

La lista de etapas no es estrictamente secuencial y algunas pueden repetirse más de una vez y pueden involucrar iteración o puede ser necesario volver a ejecutar etapas.

Además de las tres fases presentadas antes se suele considerar dos actividades previas a cualquier revisión: la identificación de su necesidad y la gestión del proyecto de la revisión. A continuación se presentan brevemente estas dos actividades y luego las actividades dentro de cada fase.

III-B1. Identificar la necesidad de una revisión: Antes de comenzar es necesario preguntarse [1, cap. 3]:

- ¿Es probable que la investigación contribuya a nuestro conocimiento sobre el tema?
- Dados los recursos disponibles (estudios primarios, equipo de trabajo, tiempo, etc.) ¿es factible la revisión?

Que una revisión sea necesaria y factible depende de varios factores. Por ejemplo, no sería necesario llevar adelante la revisión si ya existe otra que aborde el mismo tema, o uno similar, y sea de buena calidad. Por otra parte, si hay demasiados estudios primarios para analizar y la cantidad de personas o el tiempo para analizarlos no son suficientes, entonces la revisión no es factible. Otro caso en el que la revisión puede no ser factible es cuando hay muy pocos estudios de buena calidad para que su síntesis sea significativa.

Se puede realizar una evaluación crítica a un estudio secundario ya existente utilizando una lista de verificación o checklist. A continuación se incluye una de ejemplo [10]:

- ¿Fue la pregunta de investigación claramente definida en términos de población, intervenciones, comparaciones, resultados y diseño de estudios?
- ¿Fue la estrategia de búsqueda adecuada y apropiada? ¿Hubo alguna restricción de lenguaje, estado o fecha de publicación?
- ¿Se realizaron acciones preventivas para minimizar sesgo y errores en el proceso de selección de estudios?
- ¿Se usó un criterio apropiado para evaluar la calidad de los estudios primarios y se realizaron acciones preventivas para minimizar sesgo y errores en el proceso de evaluación de la calidad?
- ¿Se realizaron acciones preventivas para minimizar sesgo y errores en el proceso de extracción de datos?
- ¿Son adecuados los detalles presentados de cada estudio primario?
- ¿Son apropiados los métodos utilizados para la síntesis de datos? ¿Hay diferencias entre los estudios evaluados? ¿Se agruparon los estudios, y sí fue así: era apropiado y significativo hacerlo?
- ¿Reflejan las conclusiones de los autores la evidencia que se revisó?

III-B2. Gestionar el proyecto de la revisión: Al comienzo, es importante considerar cómo se gestionará el proyecto de revisión en su conjunto [1, sec. 4.2]. Las actividades de gestión incluyen:

- Organizar el desarrollo y validación del protocolo de revisión,
- Especificar el cronograma de la revisión,

- Asignar las tareas a los miembros del equipo,
- Decidir qué herramientas usar.

III-C. Planificar la Revisión

III-C1. Especificar las preguntas de investigación: La especificación de las preguntas de investigación es una actividad crítica ya que sientan la base para: decidir qué estudios primarios incluir o excluir, decidir qué datos deben extraerse y cómo sintetizar los datos para contestar las preguntas en el reporte final. Los factores que motivan las preguntas deben ser completamente explicados.

De acuerdo a Kitchenham [9], es conveniente crear las preguntas correctas. En general una pregunta correcta es aquella que:

- Es significativa e importante tanto para investigadores como para profesionales.
- Dará lugar a cambios o incrementará la confianza en las prácticas actuales de la ingeniería de software.
- Identificará las discrepancias entre las creencias comúnmente aceptadas y la realidad.

Kitchenham et al. [1, sec. 4.3] indican que las preguntas de investigación se formulan de una de dos maneras:

- Como una comparación cuantitativa de dos (o más) tecnologías para determinar cuál es más efectiva o eficiente que otras dentro de un contexto específico.
- Como una evaluación cualitativa de una tecnología, enfoque o procedimiento específico utilizado en la investigación de ingeniería de software, con respecto a los beneficios, riesgos, valor, impacto o algún otro aspecto de su adopción.

Aspectos a considerar al crear las preguntas de investigación incluyen [9], [7]:

- La población en la cual se recolecta la evidencia. Por ejemplo el grupo de programadores que probaron un nuevo método para diseñar software.
- La intervención aplicada en el estudio empírico. Por ejemplo, el nuevo método para diseñar software.
- La comparación contra la cual se compara la intervención. Por ejemplo, el método antiguo que los programadores utilizaban para diseñar software.
- Los resultados. Además de ser estadísticamente satisfactorios deben ser significativos desde el punto de vista práctico. Por ejemplo, seguramente se pueda obviar en los resultados que el nuevo método de diseño requiere el uso de hojas oficio en lugar de A4.
- El contexto del estudio. En general es una extensión de la población, puede incluir si fue realizado en la industria o la academia, en qué segmento de la industria o por ejemplo, los incentivos otorgados a los sujetos.
- Los diseños experimentales a incluir en las preguntas de investigación. Quizás se requiera restringir a revisar sólo estudios primarios correspondientes a casos de estudio, por ejemplo.

Recomendaciones [11], [12]

- Es probable que sea necesario revisar las preguntas durante el desarrollo del protocolo, a medida que aumente el entendimiento del problema.
- Quizás realizar un estudio de mapeo previo a la revisión sirva de ayuda para determinar el alcance de las preguntas de investigación.
- Es conveniente evaluar que los motores de búsqueda sean capaces de ejecutar las cadenas de búsqueda.

III-C2. Desarrollar el protocolo de revisión: Un protocolo de revisión especifica los métodos que se utilizarán para llevar a cabo una revisión sistemática específica y su definición de antemano puede ayudar a reducir la probabilidad de sesgo del investigador al limitar la influencia de las expectativas del investigador en, por ejemplo, la selección de estudios primarios o la síntesis de los resultados; puede ser evaluado por otros investigadores antes de realizar la revisión; y puede formar la base de las secciones de introducción y método del informe de la revisión.

Los componentes de un protocolo incluyen todos los elementos de la revisión más alguna información adicional de planificación:

- Antecedentes y justificación de la revisión.
- Preguntas de investigación.
- Estrategia de búsqueda para estudios primarios. Debe incluir los términos de búsqueda y recursos donde se realizará la búsqueda. Los recursos incluyen librerías digitales, revistas específicas, y actas de congresos.
- Criterios de selección de estudios. Usados para determinar cuáles estudios son incluidos o excluidos de la revisión sistemática.
- Procedimientos de selección de estudios. El protocolo debe describir cómo se aplicará el criterio de selección, por ejemplo, cuántos asesores evaluarán cada estudio primario, y cómo se resolverán los desacuerdos entre los evaluadores.
- Procedimientos y listas de verificación para evaluar la calidad.
- Estrategia de extracción de datos.
- Método o técnica de síntesis de los datos.
- Limitaciones
- Estrategia de difusión.
- Calendario del proyecto de revisión.

En la Tabla I se incluye una posible estructura para el protocolo [1].

Recomendaciones [11], [6]

- Conviene que todos los miembros de la revisión sistemática participen activamente en el desarrollo del protocolo de revisión.
- Resulta esencial realizar una prueba piloto del protocolo de investigación. Esto permite encontrar errores

Tabla I
PLANTILLA PARA UNA SLR SEGÚN [1]

1. Change Record

This should be a list or table summarizing the main updates and changes embodied in each version of the protocol and (where appropriate), the reasons for these.

2. Background

- explain why there is a need for a study on this topic
- specify the main research question being addressed by this study
- specify any additional research questions that will be addressed
- if extending previous research on the topic, explain why a new study is needed

3. Search Process

- specify and justify basic strategy: manual search, automated search, or mixed
- for automated searches, specify search terms and compounds of these and record results of any prototyping of the search strings
- for automated searches, identify resources to be used (specifying the digital libraries and search engines)
- for manual searches, identify the journals and conferences to be searched
- specify the time period to be covered by the review and any reasons for your choice
- identify any ancillary search procedures, for example, asking leading researchers or research groups, or accessing their web sites; or checking reference lists of primary studies
- specify how the search process is to be evaluated (for example, against a known subset of papers; or against the results from a previous systematic review)

4. Primary Study Selection Process

- identify the inclusion criteria for primary studies
- identify the exclusion criteria
- define how selection will be undertaken (roles of reviewers)
- define how agreement among reviewers will be evaluated
- define how any differences between reviewers will be resolved

5. Study Quality Assessment Process

- specify the quality checklists to be used
- specify how the checklist will be evaluated (if a new checklist has been developed)
- define how agreement among data extractors will be evaluated
- define how any differences between data extractors will be resolved
- identify the procedures to use for applying the checklists, such as details inclusion/exclusion, partitioning the primary studies during aggregation or meta-analysis, and explaining the results of primary studies

6. Data Extraction Process

- design data extraction form (and check via a dry run)
- specify the strategy for extracting and recording the data (for example, paper form, on-line. Form or database)
- identify how the data extraction process is to be undertaken and validated, particularly any data that require numerical calculations, or are subjective

7. Data Synthesis Process

- specify the form of analysis/synthesis to be used (for example, narrative, tabulation, meta-analysis)
- discuss how the synthesis will be validated

8. Study Limitations

- assess the threats to validity (construct, internal, external), particularly constraints on the search process and deviations from standard practice
- specify residual validity issues including potential conflicts of interest that are inherent in the context of the study, rather than arising from the plan

9. Reporting

- identify target audience, relationship to other studies, planned publications, authors of the publications
- agree in advance who will be included in the list of authors and whose assistance will be reported in the acknowledgements section.

10. Schedule

Provide time estimates for all of the major steps.

en los procedimientos de recolección y agregación de los datos. Es posible que surjan consideraciones que hagan cambiar la metodología que se pretendía usar para contestar las preguntas de investigación.

- Es probable que el protocolo sea modificado a lo largo del proceso de la revisión, por tanto se deben documentar y justificar todos los cambios realizados.

III-C3. Evaluar el protocolo de revisión: Ya que el protocolo es un elemento crítico de cualquier revisión sistemática, los investigadores deberían acordar un procedimiento para su evaluación, donde lo ideal sería tener un revisor independiente. De otra forma, se pueden utilizar a modo de checklist los criterios utilizados para la identificación de la revisión. Además, se puede chequear la consistencia interna del protocolo para confirmar que: las cadenas de búsqueda fueron derivadas correctamente; los datos a extraer y el procedimiento de análisis de los datos responderán adecuadamente a las preguntas de investigación.

Recomendaciones [11]

- Para la extracción de datos conviene contar con un documento breve que contenga únicamente la definición de los datos y las guías de extracción de datos del protocolo.
- Es necesario que se acuerde un proceso de validación separado de las pruebas piloto del protocolo. Sería ideal que este proceso fuera realizado por revisores externos.

III-D. Realizar la Revisión

III-D1. Identificar la investigación: En esta etapa se establece y utiliza una estrategia de búsqueda para obtener una lista de todas las publicaciones relevantes para las preguntas de investigación. La estrategia de búsqueda tiene que ser la definida en el protocolo de la revisión y debe ser reportada luego de forma transparente y replicable.

Es bueno descomponer las preguntas de investigación e identificar cadenas iniciales de búsqueda de acuerdo a población, intervención, comparación, resultados, contexto y diseño del estudio. Además conviene crear listas de sinónimos, abreviaciones y ortografías alternativas.

Otra alternativa, a veces utilizada como complemento, es el uso de la referencias en la búsqueda de nuevos estudios. Se parte de un estudio primario relevante encontrado previamente y se utilizan sus referencias para encontrar nuevos estudios relevantes (*backward snowballing*) o se buscan qué estudios lo ubican en sus referencias (*forward snowballing*) [7].

Al crear la estrategia de búsqueda se debe perseguir el equilibrio entre intentar conseguir todos los estudios relevantes y no recuperar demasiados falsos positivos que deban ser excluidos a mano. Dos criterios clave para evaluar la integridad de una búsqueda automatizada son el *recall* y la precisión

[1, sec. 5.1]. El *recall* de una búsqueda es la proporción (o porcentaje) de todos los estudios relevantes encontrados por la búsqueda. La precisión de una búsqueda es la proporción (o porcentaje) de los estudios encontrados que son relevantes para las preguntas de investigación que se abordan en una revisión.

En general, la elaboración de una estrategia de búsqueda es un proceso iterativo, un refinamiento basado en cierta determinación del nivel de integridad alcanzado. Esto se puede hacer, por ejemplo, teniendo un conjunto de artículos de antemano que la búsqueda tiene que recuperar [1, sec. 5.2] de forma de validar la estrategia. Este conjunto puede ser producto de búsquedas exploratorias previas o proporcionado por algún experto.

Sobre este tema:

- La definición de la estrategia de búsqueda es un tema no menor, sobretodo si se quiere evaluar de alguna forma qué tan buena es. En [13], Zhang et al. proponen una técnica sistemática de aproximación a una búsqueda casi perfecta utilizando una sub-búsqueda manual e iteraciones.

El *sesgo en la publicación* refiere a que es más probable que sean publicados estudios con resultados, en algún sentido, positivos a estudios que presentan resultados negativos. Para evitar este sesgo se puede consultar la literatura gris, como por ejemplo reportes técnicos, publicaciones rechazadas o trabajos en progreso.

Recomendaciones [11]

- Existen diferentes estrategias de búsqueda asociadas a diferentes tipos de criterios de búsqueda. Se debe elegir y justificar la estrategia de búsqueda más apropiada para las preguntas de investigación a responder.
- Se deben utilizar diferentes fuentes electrónicas; seguramente no exista una única fuente que pueda proporcionar todos los estudios primarios.
- Actualmente, los motores de búsqueda de ingeniería de software no están diseñados para soportar las revisiones sistemáticas de literatura. Así es que muchas veces las búsquedas quedan diseñadas dependientes de la fuente.
- Las búsquedas deben aplicarse sobre los mismos campos (esto es: título, resumen, etc) y utilizando los mismos filtros para todos los artículos en todas las fuentes. De todas maneras, la cadena de búsqueda debe adaptarse para poder ejecutarse en las distintas bibliotecas digitales.

III-D2. Seleccionar los estudios primarios: Una vez que se han obtenido los estudios primarios potenciales es necesario

evaluar cabalmente su relevancia. Para esto se utilizan criterios de inclusión y exclusión, los primeros definen que estudios se deben incluir como relevantes mientras que los últimos se aplican sobre los estudios seleccionados o sobre la lista inicial para remover estudios irrelevantes. Los criterios de selección de estudios pretenden identificar los estudios primarios que proporcionan evidencia directa acerca de las preguntas de investigación. A fin de reducir la probabilidad de sesgo, los criterios de selección deben ser decididos durante la definición de protocolo, aunque tal vez puedan ser refinados durante el proceso de búsqueda. Los criterios de inclusión y exclusión deberían estar basados en las preguntas de investigación; y probado o validados para asegurar que son interpretados de forma confiable y que clasifican los estudios correctamente.

Recomendaciones [11]

- En general, los resúmenes en las tecnologías de la información e ingeniería de software son demasiado pobres para poder seleccionar estudios primarios. Así que es recomendable revisar también las conclusiones.
- Para SLRs, estudios de mapeo y revisiones terciarias, en general, se recomienda tener por lo menos dos revisores. Cuando un equipo de revisores realiza la selección de estudios, se sugiere validar los resultados del proceso de selección por dos (o más) miembros del equipo que aplican los criterios de inclusión y exclusión de forma independiente en un subconjunto de los artículos recuperados en la búsqueda y verifican su nivel de acuerdo. Esta validación se realiza, por ejemplo, mediante la realización de un análisis kappa [1, sec. 5.2].

III-D3. Evaluar la calidad de los estudios: El objetivo de esta etapa es analizar y evaluar la calidad de cada estudio seleccionado a fin de determinar su inclusión o no en el proceso de extracción de datos y reporte de resultados de la revisión. En general, el propósito de la evaluación de la calidad es asegurar que los hallazgos de un estudio son relevantes y no sesgados.

Ante la dificultad inicial de contar con un acuerdo en la definición de calidad de un estudio, Kitchenham [9] sugiere que la calidad se refiere a la medida en que el estudio minimiza el sesgo y maximiza tanto la validez interna como la externa. A continuación se detallan estos conceptos:

- Sesgo (o error sistemático): Tendencia a producir resultados que se apartan sistemáticamente de los resultados verdaderos.
- Validez interna: El grado en que el diseño y la realización de un estudio tienden a prevenir el error sistemático.
- Validez externa: El grado en que los efectos observados en el estudio son aplicables fuera del estudio en sí.

En general, se llama instrumentos de calidad a las evaluaciones detalladas de calidad; las cuales son, generalmente,

listas de verificación de los posibles factores de sesgo que es necesario evaluar en cada estudio. La mayoría de las listas de verificación de calidad incluyen preguntas para evaluar el grado en que los artículos consideran el sesgo y la validez. También, es posible considerar: ítems genéricos relacionados a características del diseño particular del estudio e ítems específicos relacionados al área de la revisión.

A modo de referencia, Kitchenham [9] incluye una lista bastante amplia de preguntas para estudios cuantitativos y otra para estudios cualitativos. También indica la alternativa de utilizar una escala de medida para cada ítem en las ocasiones en las cuales una respuesta de Si/No puede resultar engañosa.

Es importante que los investigadores no solo definan en el protocolo el instrumento de calidad sino que también especifiquen cómo será usada la información de calidad. Las opciones son las siguientes:

- Para asistir en la selección de estudios primarios. En este caso, la información de calidad se utiliza para construir criterios de inclusión/exclusión detallados.
- Pueden llevarse a cabo análisis con o sin estudios de baja calidad para determinar el impacto de dichos estudios en los resultados finales.
- Una de las preguntas de investigación que aborda una revisión puede centrarse en las tendencias en la calidad (o su evolución) de los estudios primarios relacionados con el tema de una revisión.

Es posible contar con ambos tipos de información de calidad en una misma revisión sistemática.

Limitaciones de la Evaluación de la Calidad [9]: A menudo el informe de los estudios primarios es pobre, por lo que quizás no sea posible determinar la forma de evaluar un criterio de calidad. Es tentador suponer que porque algo no se encuentra reportado, no se hizo. Esta suposición puede ser incorrecta. Los investigadores deberían intentar obtener más información por parte de los autores del estudio.

Es posible identificar un análisis estadístico inadecuado o inapropiado, pero sin acceso a los datos originales no es posible corregirlo. Es común que los datos sean confidenciales y que no puedan estar disponibles. En algunos casos, los involucrados pueden negarse a compartir sus datos con otros investigadores debido a que desean continuar publicando sobre el tema.

La validación es un elemento importante para mantener la confianza en los procedimientos y resultados de una revisión [1, cap. 7]. Si la evaluación de la calidad es llevada a cabo por un equipo de investigadores, entonces dos o más de ellos pueden puntuar cada estudio seguido por un proceso de resolución. Dicho proceso puede ser o bien llegar a un consenso entre los investigadores, o realizar un promedio entre los puntajes otorgados.

Recomendaciones [11]

- Existe un énfasis en la necesidad de evaluar la calidad de los estudios primarios. Sin embargo, esto

depende del tipo de revisión sistemática de literatura que se lleve a cabo. Por ejemplo, es común que para algunos tipos de revisión ya es un criterio de calidad suficiente que los estudios primarios hayan sido publicados en alguna revista académica.

- Es importante asegurar que la evaluación de calidad sea utilizada en las actividades siguientes de agregación y análisis de datos.

III-D4. Extraer datos: El objetivo de esta etapa es el diseño de formularios de extracción de datos para que los investigadores puedan registrar adecuadamente la información que obtienen de los estudios primarios. A fin de reducir posibles sesgos, se recomienda definir los formularios de extracción de datos y realizar pruebas piloto durante la definición del protocolo de la revisión.

Los formularios de extracción de datos deben ser diseñados para recolectar toda la información necesaria para contestar las preguntas de investigación así como los criterios de calidad de los estudios. Si los criterios de calidad van a ser utilizados para identificar criterios de inclusión/exclusión, entonces se requieren formularios separados (ya que la información debe ser recolectada previo a la extracción de datos).

En general un formulario de extracción de datos debería incluir las siguientes partes (secciones) [6]: información de la extracción (responsable y fecha de la extracción, responsable de verificar), información general de los estudios (identificador del estudio, título y detalles de publicación), preguntas para responder las preguntas de investigación, preguntas para evaluar la calidad de los estudios y resumen de los datos.

Es importante no incluir en la síntesis de una revisión sistemática múltiples publicaciones de los estudios, ya que informes duplicados podrían afectar seriamente el sesgo de cualquier resultado. Cuando hay publicaciones duplicadas conviene usar la más completa.

Recomendaciones [11]

- Cuando sea necesario revisar un gran número de artículos puede resultar útil que un lector extraiga los datos y otro los valide.
- Si es necesario manipular datos se debe describir y justificar las manipulaciones realizadas.
- En caso de que los datos de un estudio no estén completos se puede escribir a los autores para conseguirlos.

III-D5. Sintetizar los datos: La síntesis de los datos involucra recolectar y resumir los resultados incluidos en los estudios primarios. En general, hay dos tipos de síntesis de datos: síntesis descriptiva (o narrativa) y síntesis cuantitativa [6].

Con el fin de lograr conclusiones confiables la síntesis debería considerar la fuerza de la evidencia, explorar la

consistencia y discutir inconsistencias. El enfoque debería estar definido en el protocolo y se determina por el tipo de las preguntas de investigación, pero también por el tipo de estudios disponibles y por la calidad de los datos.

Sin importar el tipo de síntesis, se debería comenzar con la creación de un resumen de los estudios incluidos. En general se presenta una tabla con detalles importantes como son el tipo, intervenciones, número y característica de los participantes, resultados, etc. También se presentan (en la misma o en otra tabla) elementos de la calidad de los estudios y riesgos de sesgo. Además, este proceso descriptivo debería ser explícitamente riguroso y ayudar a concluir que los estudios son similares y confiables para su síntesis [6].

La información extraída sobre los estudios (por ejemplo: intervención, población, contexto, tamaño de la muestra, resultados, calidad del estudio) debería presentarse tabulada de una manera consistente con las preguntas de la revisión. Las tablas deberían estar estructuradas para resaltar similitudes y diferencias entre los resultados de los estudios.

Es importante identificar si los resultados de los estudios son consistentes (en un sentido de síntesis) unos con otros (o sea, homogéneos) o inconsistentes (o sea, heterogéneos). Los resultados quizás puedan ser tabulados para mostrar el impacto de potenciales fuentes de heterogeneidad (por ejemplo: tipos de estudio, calidad del estudio o tamaño de la muestra).

Kitchenham sugiere realizar un análisis de sensibilidad para determinar si los estudios de baja calidad tienen impacto significativo sobre los resultados de la síntesis. El análisis de sensibilidad también puede realizarse sobre diferentes subconjuntos de los estudios primarios para determinar la robustez de los resultados.

Sobre este tema:

- La síntesis de los datos es un tema con bastantes componentes estadísticos y matemáticos. Inicialmente se pensaba que no iba ser posible realizar síntesis cuantitativas en ingeniería de software (debido a la calidad de los datos) y solo se explicaba en detalle la síntesis narrativa [9]. En [14] plantean una revisión terciaria estudiando los tipos de síntesis realizadas hasta el momento y se percibe cierto incremento en la formalidad al realizar síntesis de datos. El estudio es muy completo y es bastante ilustrativo, sobre todo porque hace una introducción bastante profunda del tema, así como una lista de los tipos de síntesis disponibles y recomendaciones sobre su uso. Para complementar este tema, se puede consultar [15] donde se presenta una visión sobre la síntesis particular de casos de estudios. Un estudio bastante más exhaustivo sobre síntesis de datos y meta-análisis puede leerse en [10].
- El libro de [1] tiene dos capítulos relacionados a síntesis muy buenos: el cap. 10 que presenta métodos de síntesis cualitativa y el cap. 11 que detalla meta-análisis (síntesis cuantitativa).

III-E. Informar sobre la revisión

La fase final de una revisión sistemática involucra redactar los resultados de la revisión y difundir los resultados a potenciales interesados.

Sobre este tema:

- Escribir artículos científicos, o material publicable en general, es difícil al principio y es necesario completar cierta curva de aprendizaje. Como siempre, la práctica hace al maestro. Un libro muy bueno para comenzar es uno publicado por Gustavii [16].

III-E1. Especificar los mecanismos de difusión: Es importante comunicar los resultados de la revisión sistemática. Por esto es que algunas guías recomiendan planificar la estrategia de difusión durante la etapa de puesta en marcha de la revisión (si existe) o cuando se prepara el protocolo de revisión sistemática. La mayoría de los académicos asumen que la difusión se trata de publicar en revistas académicas o presentar en conferencias. Sin embargo, si se quiere que los resultados tengan influencia sobre la práctica profesional seguramente sean necesarios otros medios de comunicación, en particular: Revistas con orientación profesional, prensa popular o especializada, folletos breves con resúmenes, posters, páginas webs, entre otros.

III-E2. Formatear el informe principal: Kitchenham [9] sugiere la estructura y contenido para informes de revisiones sistemáticas que muestra la Tabla II.

III-E3. Evaluar el informe: Los informes de la revisión sistemática suelen ser evaluados en el marco de su difusión: un artículo en una revista académica será revisado previo a su publicación, expertos revisarán una tesis de doctorado. Si el protocolo fue revisado por un grupo de expertos previo a la ejecución de la revisión entonces se recomienda que el mismo grupo revise el informe final. El proceso de evaluación puede utilizar las listas de verificación de calidad para revisión sistemáticas similares a las vistas en la sección.

IV. ESTUDIOS DE MAPEO

Los estudios de mapeo (o de alcance) son un tipo de revisión sistemática de literatura, pero a diferencia de las revisiones sistemáticas convencionales su propósito es encontrar y clasificar estudios primarios dentro de un tópico específico [17].

IV-A. Diferencias entre un estudio de mapeo y una SLR

La diferencia principal es que una revisión sistemática convencional intenta agregar los estudios primarios en términos de los resultados de la investigación e investiga si son consistentes o contradictorios. En cambio, un estudio de mapeo aspira solamente a clasificar la literatura relevante y a clasificar los estudios con respecto a categorías definidas. La Tabla III presenta las diferencias entre los estudios de mapeo y las revisiones sistemáticas convencionales [18].

IV-B. Esquemas de clasificación

Uno de los principales focos al realizar estudios de mapeo es la creación de un esquema de clasificación. Aunque si bien pueden tenerse ideas iniciales, su construcción finaliza cuando se dispone de los datos de los estudios relevantes encontrados.

En un estudio de mapeo en general se trabaja con uno o más esquemas de clasificación a fin de categorizar los estudios primarios encontrados, por ejemplo, identificando a qué país pertenece o estableciendo de qué tópico trata a partir de una lista de temas. Esto puede implicar una sola dimensión (por ejemplo, clasificar según país de publicación) o varias dimensiones (por ejemplo, clasificar según país, pero también según tema y fuente de publicación).

Para elegir un esquema de clasificación se puede consultar investigación previa o bibliografía relevante antes de comenzar el estudio de mapeo. En general, es una buena práctica validar el esquema revisando un subconjunto de las publicaciones que serán cubiertas mediante el estudio previo a realizarlo.

Si es necesario la construcción de un esquema de clasificación es posible utilizar, por ejemplo, un *keywording* de los resúmenes [19], esto implica la extracción de los términos clave de los resúmenes de los artículos encontrados. Esto puede ser bueno para algunos casos, pero en general conviene tener una noción básica de aspectos de clasificación y construcción de vocabularios controlados (que es el concepto real asociado a los esquemas de clasificación). Para esto se pueden consultar la guía de Zeng [20].

Al comenzar a trabajar con estudios de mapeo es difícil imaginar que esquemas de clasificación pueden resultar útiles. Algunos de las siguientes ideas pueden ayudar:

- Fechas de publicación — Una buena dimensión a estudiar es la fecha de publicación, en general permite analizar la evolución de las demás dimensiones a través del tiempo. En general se extrae el año solamente, pero para estudios particulares podría utilizarse el mes también.
- País de publicación — Muchas veces es de interés clasificar los estudios primarios por país de publicación.
- Fuente — También puede ser de interés clasificar los artículos por fuente de publicación.
- Métodos de investigación — Un aspecto muy interesante al evaluar artículos es determinar que método de investigación utilizan los autores. Pueden encontrarse distintas maneras de realizar esta clasificación, pero está bastante generalizado el uso del esquema planteado por Wieringa [21]. Para cada método de investigación Wieringa presenta criterios de evaluación que pueden ser utilizados, por ejemplo, para crear criterios de evaluación de la calidad. La Tabla IV muestra un resumen de la clasificación de Wieringa.

Sobre este tema:

- En [19] brindan una explicación bastante completa del proceso de creación de esquemas de clasificación que utilizaron en un caso de ejemplo.

Tabla II
ÍNDICE SUGERIDO PARA INFORME DE UNA REVISIÓN SISTEMÁTICA DE LITERATURA.

Sección	Subsección	Alcance / Comentarios
Título		Debería ser corto pero informativo, así como basado en las preguntas a responder por la revisión. Para revistas académicas, en general se incluye que se trata de una revisión sistemática.
Autoría		
Resumen Ejecutivo o Estructurado	Contexto	La importancia de las preguntas de investigación que aborda la revisión.
	Objetivos	Las preguntas contestadas por la revisión sistemática.
	Métodos	Fuentes de datos, selección de estudios, evaluación de calidad y extracción de datos.
	Resultados	Principales hallazgos incluyendo meta-análisis y análisis de sensibilidad.
	Conclusiones	Implicancias a la práctica y futura investigación.
Antecedentes		Justificación de la necesidad de la revisión. Resumen de revisiones previas. Descripción de la técnica de ingeniería de software que está siendo investigada así como su importancia potencial.
Preguntas de la revisión		Se debe especificar cada una de las preguntas de la revisión.
Métodos de la revisión	Fuentes de datos y estrategia de búsqueda	Estos datos se deberían basar en el protocolo original incluyendo información de los cambios que haya sufrido.
	Selección de estudios	
	Evaluación de calidad de los estudios	
	Extracción de datos	
	Síntesis de los datos	
Estudios Incluidos y Excluidos		Criterios de inclusión y exclusión. Lista de los estudios excluidos, así como las razones para la exclusión. Los criterios de inclusión y exclusión de los estudios a veces son representados como un diagrama de flujo, ya que los estudios pueden ser excluidos en distintas etapas de la revisión.
Resultados	Hallazgos	Descripción de los estudios primarios. Resultados de cualquier resumen cuantitativo. Detalles de cualquier meta-análisis. Se debería incluir resúmenes no cuantitativos de los estudios de forma de presentarlos en forma tabular. Los resúmenes cuantitativos se deberían presentar en tablas y gráficos.
	Análisis de sensibilidad	
Discusión	Hallazgos principales	Esto debe corresponder con los hallazgos presentados en la sección de resultados.
	Fortalezas y debilidades	Fortalezas y debilidades de la evidencia incluida en la revisión. Relación a otras revisiones, considerando particularmente cualquier diferencia en calidad y resultados. Una discusión de la validez de la evidencia considerando el sesgo en la revisión sistemática permite que el lector pueda evaluar la confianza que puede depositar en la evidencia recolectada.
	Significado de los hallazgos	Dirección y magnitud de los efectos observados en los estudios resumidos. Aplicabilidad (generalización) de los hallazgos. Dejar en claro en qué medida los resultados implican causalidad discutiendo el nivel de evidencia. Discutir todos los beneficios, efectos adversos y riesgos. Discutir variaciones en los efectos y sus razones.
Conclusiones	Recomendaciones	Implicaciones prácticas para el desarrollo de software. Preguntas sin responder e implicaciones para investigación futura.
Agradecimientos		Todas las personas que colaboraron en la investigación pero no califican en los criterios de autoría.
Conflictos de intereses		Se debe declarar cualquier interés secundario por parte de los investigadores (por ejemplo: un interés financiero en la tecnología que está siendo evaluada).
Referencias y Anexos		Los anexos pueden ser utilizados para: listar estudios incluidos y excluidos, documentar detalles de la estrategia de búsqueda, y listar datos sin pulir de los estudios incluidos.

- En [17] se presenta una comparación de varios estudios de mapeo y se comentan características de este tipo de revisiones.

IV-C. Presentación de resultados

La presentación de resultados de los estudios de mapeo es un aspecto significativo. Mientras que en las revisiones sistemáticas de literatura los resultados, en general, se pueden sintetizar de alguna forma (cualitativa o cuantitativa), en los estudios de mapeo esto puede ser que no resulte tan fácil de realizar. Por este motivo, es posible que se opte por tablas o visualizaciones gráficas de los resultados.

Como el propósito de un estudio de mapeo está fuertemente ligado a la clasificación de los estudios primarios es frecuente ver gráficos donde se presenta la distribución de las publicaciones para una dimensión (una categoría, se pueden utilizar

gráficos de torta u otros) o varias dimensiones (cruzando dos o tres categorías, por ejemplo en un gráfico de burbujas). La figura 1 presenta ejemplos de gráficos incluidos en el mapeo sobre gamificación en ingeniería de software realizado por Pedreira et al. [22].

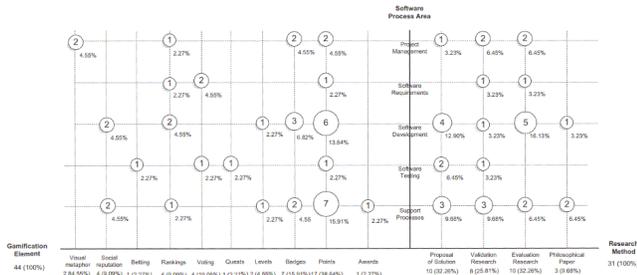
V. REVISIONES TERCIARIAS

En un dominio en el cual ya exista un número de revisiones sistemáticas puede ser posible conducir una revisión terciaria (una revisión sistemática de revisiones sistemáticas) para contestar preguntas más amplias. Una revisión terciaria utiliza exactamente la misma metodología que las revisiones sistemáticas de literatura estándares.

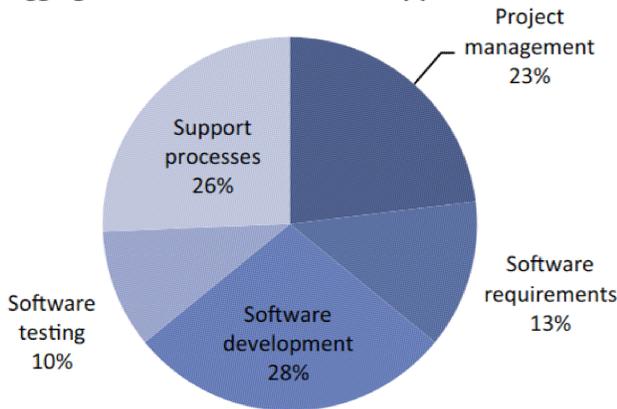
Sobre este tema:

Tabla III
RESUMEN DE DIFERENCIAS ENTRE LAS REVISIONES SISTEMÁTICAS Y LOS ESTUDIOS DE MAPEO.

Proceso de la SLR	Estudio de Mapeo	SLR
Preguntas de investigación	Generales - relacionadas a las tendencias en investigación. Qué subtópicos se manejan, qué investigadores se manejan, cuánta actividad, qué tipos de estudios, etc.	Específicas - relacionadas a los resultados de los estudios empíricos. De la forma: ¿es la tecnología/método A mejor que el B?
Proceso de búsqueda	Definido por el área temática.	Definido por las preguntas de investigación.
Estrategia de búsqueda	No muy estricta si sólo son de interés las tendencias en investigación.	Extremadamente estricta - todos los estudios relevantes deben ser encontrados.
Evaluación de calidad	No es esencial.	Es importante para asegurar que los resultados se basan en la evidencia de mejor calidad.
Resultados	Conjunto de artículos relacionados a un área temática y su clasificación según varias categorías.	Respuestas a preguntas de investigación específicas, posiblemente con calificadores (por ejemplo: los resultados aplican únicamente a novatos).



Aggregated distribution of studies by process area



Distribution of primary studies by year

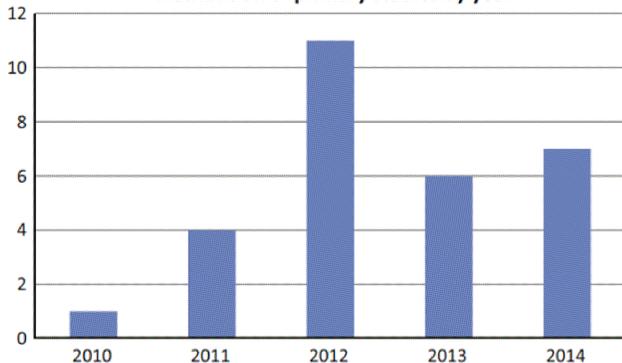


Figura 1. Ejemplo de gráficos para presentar resultados obtenidos en un estudio de mapeo.

Tabla IV

ESQUEMA DE CLASIFICACIÓN SEGÚN MÉTODOS DE INVESTIGACIÓN [21].

Categoría	Descripción
<i>Evaluation research</i>	Se investiga la aplicación práctica de una técnica. Se incluye cómo se implementa la técnica y cuáles son sus consecuencias (ventajas y desventajas). En general la técnica no es novedosa pero su aplicación práctica sí.
<i>Proposal of solution</i>	Se propone una solución a un problema y se argumenta su relevancia sin una validación en toda regla. La solución propuesta debe ser novedosa o por lo menos una extensión significativa de una técnica ya existente.
<i>Validation research</i>	Este tipo de artículos investigan las propiedades de una solución propuesta pero que aún no ha sido implementada. Posibles métodos para realizar esta validación pueden ser: experimentos, simulaciones, prototipos, análisis matemático, etc.
<i>Philosophical papers</i>	Explican una manera nueva de ver las cosas, un nuevo marco conceptual, etc.
<i>Opinion papers</i>	Contienen la opinión del autor sobre si algo está mal o bien, o cómo se hacen las cosas, etc.
<i>Personal experience papers</i>	En este tipo de artículos se explican experiencias personales. Su foco es el qué y no el porqué. La evidencia presentada en este tipo de artículos en general es anecdótica.

- A modo de ejemplo o para entender aspectos avanzados sobre revisiones sistemáticas en ingeniería de software, se pueden consultar revisiones terciarias en ese tema en [23]–[26].

VI. REVISIONES RÁPIDAS (RAPID REVIEWS)

Una evaluación realizada en medicina mostró que los responsables políticos encargados de la toma de decisiones (*policy makers*) quieren que las revisiones: respondan la pregunta correcta, sean oportunas (se completen en días o semanas

en lugar de meses o años) y sean creíbles (esto es, precisas y reproducibles) [27]. En ese sentido las revisiones rápidas buscan cumplir con esas expectativas. Fueron trasladadas recientemente a Ing. de Software por Cartaxo et al. [28]. Tienen como propósito entregar evidencia con tiempos razonables.

La estrategia más común para completar una revisión rápida dentro de un marco de tiempo limitado es reducir su alcance, esto se hace limitando las preguntas o haciendo que sean más específicas, limitando estudios primarios (por ej., utilizando un sólo motor/biblioteca o acotando la búsqueda por fecha, lenguaje o área geográfica) o limitando los resultados (presentando los resultados sin evaluación de calidad o sin síntesis formal) [27], [28]. Se considera crítico contar con la participación temprana y continua del solicitante y de cualquier otro involucrado relevante para comprender sus necesidades, el uso previsto de los resultados de la revisión, así como para realizar validaciones a lo largo del proceso [27]. Tanto la selección de estudios como la extracción de datos es, en general, realizada por un único revisor [27], [28].

Las siguientes son algunas recomendaciones realizadas por Tricco et al. [27].

- Usar un equipo experiente en ejecutar revisiones sistemáticas para conducir una revisión rápida.
- Desarrollar un protocolo que guíe la revisión y sirva para registrar todas las decisiones.
- Consultar al solicitante sobre el tipo de reporte a generar como resultado de la revisión.
- Utilizar por lo menos dos motores o bibliotecas.
- Si hay tiempo y recursos, hacer la selección con dos revisores.

Como resultado de las reducciones en la metodología, las revisiones rápidas presentan más limitaciones que las SLRs. Una manera inteligente de lidiar con estas limitaciones es dimensionarlas antes de definir qué actividades serán realizadas, por ej. consultando una tabla similar a la tabla V (basada en [28], [29]).

Algo que podría ser visto como una limitación son los tipos de preguntas que es posible contestar con este tipo de revisión. Cartaxo et al. [28] consideran que “para proporcionar respuestas útiles, uno tiene que hacer las preguntas correctas”, en ese sentido recomiendan las revisiones rápidas especialmente para preguntas exploratorias (por ej., ¿Cuáles son las estrategias para mejorar la motivación del equipo de desarrollo de software? o ¿Cuáles son las estrategias para introducir prácticas ágiles en un equipo de desarrollo ad-hoc?) y para las preguntas motivacionales (por ej., ¿Cuáles son los beneficios de las pruebas unitarias sobre la calidad del software?, es una pregunta que podría ser útil para convencer a alguien sobre la importancia de adoptar este tipo de pruebas).

VII. TRADUCCIÓN DE CONOCIMIENTO

La traducción de conocimiento (KT, por sus siglas en inglés) y la difusión son dos etapas que suceden a la ejecución de un estudio secundario y que permiten, como muestra la figura 2, transformar los datos obtenidos en forma de evidencia

Tabla V
POSIBLES SESGOS POR ACTIVIDAD DE LAS RR

Posibles actividades	Sesgos potenciales/Problemas
Un sólo revisor durante la selección de estudios	Sesgo de selección (el investigador no identifica todos los estudios primarios disponibles sobre un tema). Menos transparencia y reproducibilidad.
Seleccionar estudios primarios sólo por el título	Sesgo de selección (el investigador no identifica todos los estudios primarios disponibles sobre un tema).
Un sólo revisor durante la extracción de datos	Errores en la extracción de datos que puedan llevar a malas interpretaciones de los resultados de los estudios primarios
Un sólo motor o biblioteca digital	Posible limitación en los estudios primarios encontrados.
No conducir evaluación de calidad	Posible limitación en la confiabilidad en la evidencia encontrada.

a conocimiento que aplican los practicantes. Ambas etapas involucran temas complejos.

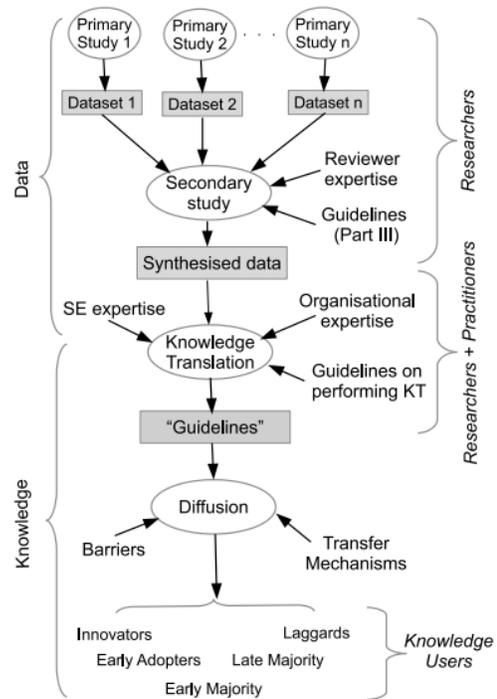


Figura 2. El camino desde los datos hasta el conocimiento [1, cap. 14]

Una definición de KT es la siguiente: *El intercambio, síntesis y aplicación éticamente sólida de conocimiento, dentro de un complejo sistema de interacciones entre investigadores y usuarios, para acelerar la captura de los beneficios de la investigación para ayudar a crear software de mejor calidad y mejorar los procesos de desarrollo de software.*

En la práctica puede involucrar producir recomendaciones a partir de lo obtenido en un estudio secundario en forma conjunta con practicantes o expertos en el área. Las recomendaciones deben transmitir un mensaje claro y deben ser tan simples como sea posible para seguir en la práctica. Se debe proporcionar una indicación de qué tan fuerte es una

recomendación (y qué se quiere decir con esto). También es importante, identificar la audiencia para la recomendación y, cuando corresponda, si la evidencia proviene de usar practicantes o estudiantes como participantes en los estudios primarios.

La difusión involucra entonces hacer llegar ese conocimiento (en forma de recomendaciones) a los practicantes para que lo utilicen. Hay todo un campo de estudio en difusión de investigación e innovaciones, cuya introducción escapa a esta guía y puede ser consultado en [1, cap. 14].

Un enfoque interesante y reciente para la traducción y difusión son los *evidence briefings* [30]. Los *evidence briefings* son un medio para transferir el conocimiento a la práctica. Consisten en un documento de una página que informa sobre los principales hallazgos de una investigación empírica. Presentan sobre todo las recomendaciones para la práctica y su contexto de aplicación, un ejemplo se presenta en la Figura 3.

The image displays a document titled "CUSTOMER COLLABORATION IN SOFTWARE PROJECTS". At the top, there are logos for ESEG, Empirical Software Engineering Group, UFPE, and Instituto Federal de Pernambuco. The document is structured as follows:

- Introduction:** "This briefing reports scientific evidence from a Rapid Review on the benefits, challenges, and strategies to establish customer collaboration in software development projects."
- FINDINGS:** A summary of 17 scientific studies, including an analysis of agile teams in Europe, a survey of agile practitioners in China, and a case study of a Norwegian company.
- BENEFITS OF CUSTOMER COLLABORATION:**
 - It drives agile behavior.
 - It permits an evolutionary model of system design and deployment.
 - It leads to greater productivity.
 - It leads to customer satisfaction.
 - It increases system quality.
 - It reduces project costs.
 - It produces systems that meet or exceed customer expected functions.
 - It is associated with projects that experienced a lesser magnitude of effort overrun.
 - It impacts on the overall success of the requirements elicitation process.
- PROBLEMS DUE TO LOW CUSTOMER COLLABORATION:**
 - Problems in Gathering and Clarifying Requirements
 - Problems in Prioritizing Requirements
 - Problems in Securing Feedback
 - Lack of Productivity
 - Business Loss
- CHALLENGES TO ESTABLISH CUSTOMER COLLABORATION:**
 - Alignment of repeatable patterns:** Interactions with customers are highly situated, which hinders the possibility to define repeatable patterns and reuse in different projects.
 - The customer representative is rarely ideal:** This affects the nature of customer collaboration and communication. For example: how much authority the customer has in making decisions, how much knowledge of the domain the customer has, where (geographically) the customer is located relative to the developers.
 - Skeptical and hyper-some skeptic customers:** don't understand agile practices such as "fail fast" and its intended benefits. A customer stated to "Forget about fail early, we don't want to fail at all!"
 - On the other extreme, customers treat agile as a buzzword and are eager to apply agile without fully understanding their own collaboration responsibilities.**
 - Lack of time commitment:** software teams realize that the customer representative's operational job may sometimes take precedence over their involvement on agile projects. Customer representative's ability to devote time for collaboration is dependent on his/her boss.
 - Absence of indirect links:** the indirect links are customer surrogates. Indirect links are less desirable due to information filtering and distortion.

Additional sections include "STRATEGIES TO ESTABLISH CUSTOMER COLLABORATION", "Who is this briefing for?", "Where the findings come from?", "What is a Rapid Review?", "What is included in this briefing?", "What is not included in this briefing?", and "To access other evidence briefings like this: citr.upr.br/seg/evidence-briefings".

Figura 3. Ejemplo de *evidence briefing*.

REFERENCIAS

[1] B. Kitchenham, D. Budgen, and P. Brereton, *Evidence-based software engineering and systematic reviews*, ser. Chapman & Hall/CRC Innovations in Software Engineering and Software Development Series. CRC Press, 2015.

[2] B. Kitchenham, T. Dyba, and M. Jorgensen, "Evidence-based software engineering," in *International Conference On Software Engineering*, 2004, pp. 273–281.

[3] R. I. Horwitz, A. Hayes-Conroy, R. Caricchio, and B. H. Singer, "From evidence based medicine to medicine based evidence," *The American journal of medicine*, vol. 130, no. 11, pp. 1246–1250, 2017.

[4] T. Dybå and T. Dingsøy, "Strength of evidence in systematic reviews in software engineering," no. June 2014, p. 178, 2008.

[5] M. Jørgensen, "A review of studies on expert estimation of software development effort," *Journal of Systems and Software*, 2004.

[6] Z. Stapić, E. G. López, A. G. Cabot, L. de Marcos Ortega, and V. Strahonja, "Performing systematic literature review in software engineering," in *CECIS 2012-23rd International Conference*, 2012.

[7] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in software engineering*. Springer Science & Business Media, 2012.

[8] S. Easterbrook, J. Singer, M.-A. Storey, and D. Damian, "Selecting empirical methods for software engineering research," in *Guide to advanced empirical software engineering*. Springer, 2008, pp. 285–311.

[9] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," Keele University, Tech. Rep., 2007.

[10] Centre for Reviews and Dissemination, *Systematic Reviews: CRD's guidance for undertaking reviews in health care*. University of York, 2009.

[11] P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil, "Lessons from applying the systematic literature review process within the software engineering domain," *Journal of Systems and Software*, 2007.

[12] J. Biolchini, P. G. Mian, A. C. C. Natali, and G. H. Travassos, "Systematic review in software engineering," *System Engineering and Computer Science Department COPPE/UFRJ, Technical Report ES*, vol. 679, no. 05, p. 45, 2005.

[13] H. Zhang, M. A. Babar, and P. Tell, "Identifying relevant studies in software engineering," *Information and Software Technology*, vol. 53, no. 6, pp. 625–637, 2011.

[14] D. S. Cruzes and T. Dybå, "Research synthesis in software engineering: A tertiary study," *Information and Software Technology*, vol. 53, no. 5, pp. 440–455, 2011.

[15] D. S. Cruzes, T. Dyba, P. Runeson, and M. Host, "Case studies synthesis: Brief experience and challenges for the future," in *2011 International Symposium on Empirical Software Engineering and Measurement*. IEEE, 2011, pp. 343–346.

[16] B. Gustavii, *How to Write and Illustrate a Scientific Paper*, 2nd ed. Cambridge University Press, 2008.

[17] D. Budgen, M. Turner, P. Brereton, and B. A. Kitchenham, "Using mapping studies in software engineering," in *PPIG*, vol. 8, 2008, pp. 195–204.

[18] B. A. Kitchenham, D. Budgen, and O. P. Brereton, "The value of mapping studies—a participant-observer case study," in *EASE*, vol. 10, 2010, pp. 25–33.

[19] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, "Systematic mapping studies in software engineering," in *Ease*, vol. 8, 2008, pp. 68–77.

[20] M. L. Zeng, *Constructing Controlled Vocabularies: A Primer*, 2005.

[21] R. Wieringa, N. Maiden, N. Mead, and C. Rolland, "Requirements engineering paper classification and evaluation criteria: a proposal and a discussion," *Requirements engineering*, vol. 11, no. 1, pp. 102–107, 2006.

[22] O. Pedreira, F. García, N. Brisaboa, and M. Piattini, "Gamification in software engineering – a systematic mapping," *Information and Software Technology*, vol. 57, pp. 157 – 168, 2015.

[23] B. Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering—a systematic literature review," *Information and software technology*, vol. 51, no. 1, pp. 7–15, 2009.

[24] B. Kitchenham, R. Pretorius, D. Budgen, O. P. Brereton, M. Turner, M. Niazi, and S. Linkman, "Systematic literature reviews in software engineering—a tertiary study," *Information and software technology*, vol. 52, no. 8, pp. 792–805, 2010.

[25] F. Q. Da SILVA, A. L. Santos, S. Soares, A. C. C. França, C. V. Monteiro, and F. F. Maciel, "Six years of systematic literature reviews in software engineering: An updated tertiary study," *Information and Software Technology*, vol. 53, no. 9, pp. 899–913, 2011.

- [26] H. Zhang and M. A. Babar, "Systematic reviews in software engineering: An empirical investigation," *Information and Software Technology*, vol. 55, no. 7, pp. 1341–1354, 2013.
- [27] A. C. Tricco, E. V. Langlois, and S. E. Straus, *Rapid Reviews to Strengthen Health Policy and Systems: A Practical Guide*, 2017, oCLC: 1050726238.
- [28] B. Cartaxo, G. Pinto, and S. Soares, "The Role of Rapid Reviews in Supporting Decision-Making in Software Engineering Practice," in *Proceedings of the 22nd International Conference on Evaluation and Assessment in Software Engineering 2018 - EASE'18*. Christchurch, New Zealand: ACM Press, 2018, pp. 24–34.
- [29] V. C. Universtiy", "Research Guides - Rapid Review Protocol," <https://guides.library.vcu.edu/rapidreview>, Dec. 2018.
- [30] B. Cartaxo, G. Pinto, E. Vieira, and S. Soares, "Evidence briefings: Towards a medium to transfer knowledge from systematic reviews to practitioners," in *International Symposium on Empirical Software Engineering and Measurement*, 2016, pp. 57:1–57:10.