

Proceso de definición de métricas y criterios de calidad en el ciclo de vida de la educación virtual accesible

Luis Fernández¹, Carmen Pagés¹, María José Rueda²

¹Departamento de Ciencias de la Computación
E.T.S. de Ingeniería Informática
Universidad de Alcalá
28871 Alcalá de Henares (Madrid)

Email: luis.fernandezs@uah.es, carmina.pages@uah.es

²DEISER SL
Calle de Valentin Beato, 22
28037 Madrid
mariajose.rueda@deiser.com

Resumen. Este trabajo describe el proceso de definición de métricas realizado durante la generación de la guía metodológica para un ciclo de vida de la educación virtual accesible del proyecto ESVIAL. Basado en los fundamentos de la teoría de la medición, se describen los criterios de revisión de las propuestas realizadas por los distintos grupos de trabajo que han elaborado los detalles de cada una de las tareas dentro de los procesos del ciclo de vida.

Palabras clave: e-learning, accesibilidad, métricas, estándares, ciclo de vida

1. Introducción

La educación virtual o *e-learning* ha recorrido ya un amplio camino de evolución al que han contribuido una gran cantidad de propuestas, técnicas y metodologías. Como toda área con cierto grado de desarrollo, ha surgido la necesidad de estandarizar algunos de los aspectos involucrados en la disciplina por lo que ya existe más de un centenar de estándares relacionados con este campo [1]. Entre ellos, podemos encontrar aquellos relacionados con la calidad y accesibilidad de la formación virtual. El proyecto ESVIAL (www.esvial.org) tiene como uno de sus objetivos el establecer un modelo de trabajo para el cumplimiento de requisitos y estándares de accesibilidad en el contexto de la formación virtual, especialmente a través de la Web. Para ello, se ha procedido a crear una guía que incluye una estructura de procesos o etapas que permita llevar a cabo un proyecto educativo virtual accesible, definiendo los distintos componentes que forman parte de dichos procesos. Esta propia estructura de procesos de la guía se ha creado para cumplir el estándar internacional ISO/IEC 19796. Este estándar establece en su primera parte las categorías de procesos que deberían considerarse para los sistemas de información y comunicación para aprendizaje, educación y formación [2].

En la guía, para cada proceso se han establecido las actividades en las que éste se descompone. El estándar ISO/IEC 19796-1 utilizado como referencia, sólo establece las categorías de procesos a considerar, y no define su descomposición, aunque la norma sí incluye algunos ejemplos de referencia. Las actividades que se han considerado para cada proceso totalizan 32, con un número de entre tres y seis actividades por cada proceso. Cada una de las actividades definidas para realizar cada proceso, se ha descompuesto en tareas, con un número de entre dos y siete. Para cada tarea, se han definido los resultados que se deberían obtener, las técnicas que se sugiere aplicar y algunas métricas y/o criterios de calidad que ayuden a evaluar el desarrollo de la tarea. En este artículo describiremos cómo se ha desarrollado una primera versión de métricas de calidad para la guía metodológica del proyecto ESVIAL. En el apartado 2 comentaremos los fundamentos de medición aplicados durante el proceso mientras que en el apartado 3

2. Fundamentos de medición

La medición se define como el proceso por el cual se asignan números o símbolos a atributos o características de entidades del mundo real de tal forma que los describa de acuerdo con reglas claramente definidas [3]. Por tanto, al definir una métrica, debemos especificar:

- La entidad sobre la que se mide: puede ser un producto tangible (por ejemplo, la plataforma o sistema), intangible (por ejemplo, un diseño) o un proceso (por ejemplo, la tutorización de alumnos).
- El atributo o característica que se mide sobre dicha entidad: por ejemplo, la facilidad de uso (atributo) de un contenido (entidad).
- Reglas claras, bien definidas y consistentes de asignación de números o símbolos para cualquier caso de entidades.

Sin estos mínimos elementos, una propuesta de métrica no alcanza el mínimo para ser considerada válida. Lo primero que toda medida debe definir claramente es su objetivo: qué entidad y qué atributo pretende caracterizar. Otra de las ideas que se pueden extraer de la definición de medición es que, antes de medir un atributo de una entidad, hay que analizar y tener muy clara la idea de lo que significa ese atributo. La exactitud de la medición se encontrará directamente relacionado con nuestro grado de conocimiento y comprensión de un atributo. Si sabemos poco de un atributo, en buena parte debido a que los expertos no han llegado a un consenso, sobre él, debemos conformarnos con medidas subjetivas y poco precisas. En cuanto a las entidades, cuanto más definidas y específicas sean, más fácil resultará medir sus atributos.

Sin embargo, no basta con definir estos elementos para que una métrica sea una propuesta consistente y válida: debería proporcionarse alguna información que permita confirmar que la métrica mide, o expresa la valoración aceptada de, la característica correspondiente. Este es un enfoque similar al adoptado por el gobierno de EEUU para orientar a los *web masters* gubernamentales a la hora de aplicar orientaciones de diseño para la usabilidad de la web (www.usability.gov) de tal manera que se cataloga cada recomendación en función de las evidencias presentadas sobre sus efectos. En [4] se resume la validación de medidas como la confirmación de

la validez de los atributos (si el atributo que nos interesa realmente está presente en la entidad que medimos), la validez de las unidades (si la unidad empleada para medir es apropiada para medir el atributo), la validez del instrumento de medida (si el modelo subyacente es válido) y la validez del protocolo de medida (si el procedimiento o protocolo de medida es apropiado). Lo cierto es que no existe una prueba definitiva que indique si una medida es válida o no; así, una medida es válida si todavía no hemos conseguido comprobar que no lo es. Pero, al menos, sí podemos analizar si los elementos de su definición están completos y son consistentes con los principios ya enunciados anteriormente. Esta filosofía es la que se ha aplicado a las propuestas de métricas generadas durante el desarrollo de la guía metodológica de ESVIAL.

3. Proceso de revisión de propuestas de métricas

La elaboración de la guía de ESVIAL se ha realizado mediante grupos de trabajo concentrados en cada uno de los procesos que han realizado la elaboración de su estructura de actividades y tareas a la vez que han propuesto una serie de técnicas y métricas de soporte. En el caso del proceso de revisión de las fichas sobre métricas propuestas, los autores han constituido un equipo de trabajo para la revisión de las mismas desde el punto de vista de los fundamentos de medición y de la descripción precisa y completa de los distintos elementos de cada propuesta. Para ello, el proceso de revisión realizado consta de dos fases. En la primera se adjuntan en cada ficha de métrica la información estructurada sobre el cumplimiento de los mínimos de definición y validez respecto de la adaptación del estándar ISO/IEC 19796-3 [5], siguiendo unos criterios detallados de revisión. Los responsables corrigen las fichas en base a dicha información y podrán marcar aquellas propuestas que consideren inviables temporal o definitivamente. En la segunda, una vez recibidas las fichas corregidas, se eliminan aquellas marcadas como inviables y también las que no cuenten con un mínimo de definición y consistencia a pesar de la corrección. Sólo las que se consideren totalmente conformes aparecerán en la primera versión de la guía ESVIAL sin que esto suponga que se descarte el trabajo posterior de mantenimiento y refinamiento de la guía para incluir otras propuestas de métricas que correspondan a las temporalmente relegadas por tener pendientes algunas correcciones en su definición o otras nuevas que surjan durante la revisión de la guía: también podrá haber cambios en las incluidas en la guía a la luz de su aplicación práctica durante el proyecto.

3.1 Adaptación del estándar ISO/IEC 19796-3 a la descripción de métricas/criterios en e-learning

Debemos tener en cuenta que el esquema de métricas incluido en el estándar ISO 19796-3 [5] está orientado a las métricas pensadas para plataformas de e-learning por lo que su aplicación a otras entidades que se quieran medir podría ser más difícil de comprender. De ahí la inclusión de unos comentarios orientativos a las categorías y elementos establecidos en el estándar. Igualmente la necesidad de obtener métricas completas y usables nos ha llevado a definir algunas categorías y elementos

opcionales del estándar como obligatorios. El resultado final del estándar adaptado y comentado a la descripción de métricas/criterios en e-learning es el siguiente:

IDENTIFICADOR / IDENTIFIER (obligatoria): Se ha asignado un identificador que combina el número de capítulo y un secuencial.

NOMBRE / NAME (obligatoria): El nombre debe ser breve e identificativo del tipo de medición. En general, no será una frase completa sino una expresión más o menos completa basada en algún sustantivo.

OBJETIVO / OBJECTIVE (obligatoria): Indicar cuál es el propósito, para que serviría usarla a efectos de controlar calidad o tener un indicador para actuar.

FUENTE / SOURCE (obligatoria, opcional en el estándar): Referir a una fuente bibliográfica o indicar si ha sido propuesta por el equipo de este proyecto Alfa, si es posible indicar el nombre del autor por si se requiere alguna aclaración posterior.

DESCRIPCIÓN / DESCRIPTION (obligatoria): Un párrafo breve que indique qué característica se mide, sobre qué (indicar un proceso o un producto o un servicio, etc.) y cómo se aplica.

DERECHOS / RIGHTS (opcional)

ALCANCE / SCOPE (opcional): Qué característica se mide, normalmente debería vincularse a una subcategoría de las características de calidad de la norma ISO 19796-3 (tabla 1), relacionadas con la plataforma de aprendizaje, no específicamente diseñadas para procesos u otros elementos como el diseño instruccional.

Tabla 1. Categorías ISO 19796-3 relacionadas con la plataforma de aprendizaje

Subcategoría	Medición
Funcionalidad	Adecuación Precisión Interoperabilidad Conformidad Seguridad
Fiabilidad	Madurez Tolerancia a errores Capacidad de recuperación
Facilidad de uso	Facilidad de comprensión Facilidad de aprendizaje Capacidad de recuperación
Eficiencia	Comportamiento en el tiempo Comportamiento de los recursos
Facilidad de mantenimiento	Facilidad de análisis Facilidad de cambio Estabilidad Facilidad de prueba
Portabilidad a otros entornos	Facilidad de adaptación Facilidad de instalación Conformidad con estándares Facilidad de reemplazo de componentes o del sistema
Adecuación educativa	Adecuación a la enseñanza Claridad Conveniencia a los objetivos Variedad Iniciativa Especialidad Actualización

TIPO / TYPE: un apartado clave para definir la métrica:

- Categoría/Category (obligatorio): Se recomienda usar una de estas categorías: Categorías, clasificación, ordenación, posición, mapeado, binario (sí/no, verdad/falso, aprobado/suspenso, etc.), descripción de un comportamiento.
- Cálculo/Calculation (obligatorio, opcional en el estándar): Indicar cómo se calcula la métrica a partir de un caso real. Aportar los elementos subjetivos u objetivos que se utilizan y los posibles cálculos o fórmulas necesarias, los criterios para clasificar en una categoría u otra, etc.
- Tipo de escala/Scale type (obligatorio): describe las unidades y símbolos que se asignan para caracterizar el atributo de la entidad real que medimos. Elegir una de las siguientes:
 - Nominal: categorías, por ejemplo como "azul", "blanco", "rojo".
 - Ordinal: categorías ordenadas, por ejemplo como "muy poco", "poco", "bastante".
 - Intervalo: categorías ordenadas que mantienen entre sí una equidistancia, por ejemplo como temperatura en grado Celsius, el paso de 13° a 14° es igual que de 14° a 15°.
 - Ratio: sobre la escala de intervalo aporta la posibilidad de proporcionalidad, por ejemplo como la longitud en centímetros: una línea de 30 cm es el doble que una de 15cm pero una temperatura de 30° no es el doble que una de 15°.
 - Absoluta: aporta un valor de referencia o inicio, como un punto cero, por ejemplo, número de palabras de un texto.
- Criterio/Criterion (obligatorio, opcional en el estándar): Debe marcarse un nivel significativo o umbral que permita interpretar resultados. Por ejemplo, menos de 0° supone la congelación del agua. Una categoría de nivel al menos A2 (AA) en accesibilidad es el mínimo para una buena plataforma.
- Recursos/Resource (obligatorio, opcional en el estándar): Cuestionario, observación directa, informe de herramienta TAW, etc.

PERIODO / PERIOD (opcional pero recomendado): Por ejemplo: una vez, continuo, al terminar cada fase, etc.

ACTORES / ACTORS (opcional pero recomendado). Por ejemplo: profesor, gestor, administrador de plataforma, etc.

ANOTACIONES / ANNOTATION (opcional): Cualquier tipo de comentario aclaratorio o complementario, especialmente referencias a los párrafos del capítulo donde aparece y referencias bibliográficas con las que pueda relacionarse (especialmente si se ha inspirado en métricas o informaciones previas en otras fuentes bibliográficas)

EXPERIENCIA o EJEMPLOS/ EXPERIENCES (obligatorio): Aquí deben incluirse las referencias a aplicaciones de la métrica y experimentos, con dos propósitos: ayudar a comprender su aplicación y utilidad mediante ejemplos o experiencias y aportar evidencias de su validez a través de sus resultados prácticos. Tanto si se ha propuesto como una nueva métrica como si se trata de una aplicación nueva o ya conocida de una ya existente convendría catalogarla, aportando las referencias necesarias, según el baremo de las categorías usadas por gobierno de EEUU para evaluar las aportaciones de usabilidad en función de las evidencias que existen sobre su validez y la opinión de un grupo de expertos (Tabla 2). Se suelen recomendar principalmente las guías con nivel al menos de 3, aunque se considera que todas

podrían ser tenidas en cuenta por los profesionales de web ya que todas se consideran que podrían aportar alguna utilidad.

Tabla 2. Categorías usadas en www.usability.gov para evaluar aportaciones de usabilidad

Categoría	Descripción
5 – Apoyado en fuertes evidencias de investigación (Strong Research Support)	Evidencias acumuladas y convincentes basadas en investigaciones Al menos un estudio formal y riguroso con validez contextual No se conocen conflictos con otros hallazgos basados en investigaciones Las opiniones de los expertos coinciden con las conclusiones de la investigación
4 – Apoyado en evidencias moderadas de investigación (Moderate Research Support)	Evidencias acumuladas basadas en investigaciones Puede haber o no conflictos con otros hallazgos basados en investigaciones Las opiniones de los expertos: tienden a coincidir con las conclusiones de la investigación y/o parece posible llegar a un consenso
3 – Apoyado en evidencias débiles de investigación (Weak Research Support)	Evidencias limitadas basadas en investigaciones Pueden existir conflictos con otros hallazgos basados en investigaciones y/o hay opiniones mixtas entre el panel de expertos
2 – Apoyado en opiniones fuertemente respaldadas por los expertos (Strong Expert Opinion Support)	No hay evidencias basadas en investigaciones Los expertos tienden a estar de acuerdo pero podría no haber un consenso total Hay múltiples opiniones favorables de expertos en libros, guías de estilo, etc. Se acepta generalmente como una "buena práctica" o refleja el "estado de la práctica"
1 - Apoyado en opiniones débilmente respaldadas por los expertos (Weak Expert Opinion Support)	No hay evidencias basadas en investigaciones Opiniones favorables de expertos limitadas o en conflicto entre sí

3.2 Criterios detallados de revisión

La primera recomendación que se ha seguido es que sólo se deben proponer métricas si existe una madurez suficiente del conocimiento o de la experiencia para poder definir con garantías los elementos que se requieren, por tanto no todos los capítulos del libro o secciones de la metodología deben necesariamente anexar métricas. La segunda recomendación es la utilización como plantilla de la adaptación del estándar ISO/IEC 19796-3 [5] a la descripción de una métrica/criterio en e-learning.

Para la validación de las métricas propuesta se han establecido unos criterios detallados, concretos y precisos partiendo de las dos recomendaciones anteriores. Cada una de las métricas propuestas será revisada siguiendo los cinco pasos siguientes:

1. Contiene la información obligatoria definida como tal en la adaptación del estándar. No se valora si la información contenida es pertinente. Es importante que el nombre de la métrica propuesta haga referencia clara al escenario de

aplicación y no dependa de ninguna otra información. En algunas casos, al incluir la sección del libro en la que la métrica se referencia, el nombre dado es incomprensible fuera de contexto, ya que el autor de la ficha sobreentiende que el escenario de aplicación es conocido. La métrica debe ser independiente del libro en cuanto a su consulta.

2. No existen métricas repetidas o con información parecida. Se indican aquellas métricas que parecen iguales o muy similares respecto al objeto de medición. En los casos que no está clara la similitud, los autores deben asegurarse de que no haya métricas repetidas. Repetidas significa que representan la misma métrica, aunque no sean exactamente iguales.
3. La métrica está referenciada en alguna sección del capítulo del libro. Aunque no es obligatorio, si la métrica es referenciada en el libro es muy conveniente indicar la sección en la que se referencia.
4. Que sea una métrica:
 - a) Mide algo sobre algo. En todos los casos debe estar claro, en la categoría Descripción de la ficha, el atributo que se mide y la entidad sobre la que se mide, además de incluir un texto explicativo.
 - b) Define escala y unidades. Se verifican algunas características de los elementos del estándar Categoría y Tipo de escala, tales como: representen el objetivo de la métrica, correspondan a valores posibles entre los indicados en la adaptación del estándar y que el valor elegido en el elemento Categoría sea único. Hay casos en que el resultado de la métrica es una lista de indicadores o de medidas binarias (sí/no), por ejemplo una lista de metadatos y si están o no rellenos. El autor de la ficha debe decidir, e indicarlo expresamente, si interesa saber si cada elemento de la lista es importante o si es importante que un porcentaje de los elementos esté completado o si lo importante es la media, u otro cálculo, sobre los valores calculados de los elementos de la lista de indicadores (medida indirecta que se crea con un cálculo sobre medidas más sencillas).
 - c) Define un cálculo. El cálculo debe ser lo suficientemente preciso: que hay que hacer para obtener cada uno de los valores definidos en el elemento Tipo de escala.
5. Que referencie experiencias de aplicación.

Se ha definido una tabla para facilitar la estandarización y análisis de los resultados de la revisión y para ayudar a los autores durante el proceso de corrección, enviándoles los comentarios y sugerencias, de forma estructurada. Los resultados de los cinco pasos de la revisión, se reflejan en las columnas de la tabla:

- Paso 1. La columna Completa contiene N cuando la ficha no contiene toda la información obligatoria. Se indica la información ausente en la columna Apartados que faltan.
- Paso 2. La columna Similar contiene el identificativo de las fichas que parecen corresponder a métricas iguales o parecidas. Explicaciones adicionales pueden aparecer en la columna Comentarios.
- Paso 3. La columna Sección de referencia contiene S si la ficha indica una referencia en algún apartado impropio de la ficha, como Identificador, o, incluso, como anexo a la propia ficha. Debe solo incluirse en la categoría Anotaciones de la ficha.

- Paso 4. En la columna Mide X (atributo) sobre Y (entidad) (Atributo/Entidad) indica lo que se deduce de la lectura de la ficha: qué mide la métrica y sobre qué lo mide. En las columnas Comentarios a Categoría y Comentarios a Tipo escala se indican recomendaciones sobre el contenido de los elementos correspondientes de la ficha. En la columna Comentarios a Cálculo se indica si el cálculo no es lo suficientemente preciso.
- Paso 5. Se indica en la columna Comentarios a Ejemplo si lo referenciado en la categoría Ejemplo de la ficha no corresponde a una aplicación o experimentación de la métrica.

4 Conclusiones

La primera versión completa de las fichas comprendía un total de 95 fichas, de las cuales solo 5 cumplían los 5 criterios definidos. Los mayores problemas encontrados han sido la falta de datos obligatorios, principalmente describir una forma de cálculo precisa, y que el ejemplo de experiencia referido en la ficha, no lo era realmente de aplicación de la métrica, sino de literatura (con casos reales o no) relacionada con el tema de la métrica. El grupo de trabajo de definición de las métricas estaba formado por siete subgrupos, cada uno con varias personas, por lo que se han encontrado similitudes en 38 fichas, debido al trabajo en paralelo realizado por los grupos. Por ello, también es importante la revisión exhaustiva de la redacción y el estilo de escritura de las descripciones de métricas, analizando las equivalencias de términos y elección de los más apropiados con un glosario para homogeneizar su uso.

De esta experiencia se deduce que el proceso de definición de métricas de calidad debe basarse en unos principios claros y en el uso de un estándar de referencia, que permitan definir formatos y reglas comunes para facilitar la estandarización de la información y la coordinación. Es indispensable un proceso sistemático de revisión centralizada en varios ciclos (modelo en espiral) para obtener resultados de calidad.

Agradecimientos. Este trabajo ha sido financiado en parte por la Comisión Europea a través del proyecto ESVI-AL del programa ALFA III.

Referencias

1. Hiler, J.R., Hoya, R. (2010) Guía de consulta de estándares de e-learning. Universidad de Alcalá. <http://www.cc.uah.es/hilera/GuiaEstandares.pdf>
2. ISO (2009) ISO/IEC 19796-1:2005 ITLET Quality management, assurance and metrics, Part 1: General approach. International Organization for Standardization.
3. Fenton, N.E. y Pfleeger, S.L. (1997) Software metrics. A rigorous and practical approach, PWS Pub.
4. Kitchenahm, B., Pfleeger, S.L. y Fenton, N.E. (1995) "Towards a framework for software measurement validation", IEEE Transactions on software engineering, vol. 21, nº 12, pp. 929-944.
5. ISO (2009) ISO/IEC 19796-3:2009, ITLET Quality management, assurance and metrics, Part 3: Reference methods and metrics.