



El problema de la representación digital de los textos: más allá de XML

Domenico Fiormonte, Università Roma Tre
(fiormont@uniroma3.it)

Desmond Schmidt, Queensland University of Technology
(desmond.schmidt@qut.edu.au)

Universidad de La Coruña, 2-3 de Julio 2012

Qué es una representación?

- Las culturas se comunican, entre ellas, a través de “mecanismos metaculturales” (Lotman 1977).
- Cada codificación/representación genera, permite y regula un determinado uso y acceso al conocimiento
- Los lenguajes de marcado traducen, representan o modelan conocimiento.
- “Podemos considerar XML un mecanismo de linearización, un método de representación del conocimiento en serie y lineal” (Buzzetti 2006).

El estándar como representación cultural

- “The control of standards is a central feature of economic life... classifications and standards are material, as well as symbolic.” (Bowker & Star)
- “Cualquier intento de crear un sistema de clasificación obligatorio, rígido y universal producirá inmediatamente como resultado, categorías residuales (...) necesitamos desarrollar la conciencia de lo que ocurre cada vez que intentamos estandarizar, es decir, que en esta creación hay quien gana y quien pierde.” (S. Leigh Star)

El efecto “Lock-in”

- “The fateful, unnerving aspect of information technology is that a particular design will occasionally happen to fill a niche and, once implemented, **turn out to be unalterable.**”
- “Lock-in removes ideas that do not fit into the winning digital representation scheme”
- “Lock-in, however, removes design options based on what is easiest to program, what is politically feasible, what is fashionable, or what is created by chance.” (Jaron Lanier 2010)

Un nuevo modelo

Una infraestructura adecuada para las humanidades debería fundarse en:

- Las estructuras contenidas en los tipos de documentos que se están representando (ej. manuscritos estratificados)
- Las necesidades del usuario (editor, estudiante, investigador, etc.)

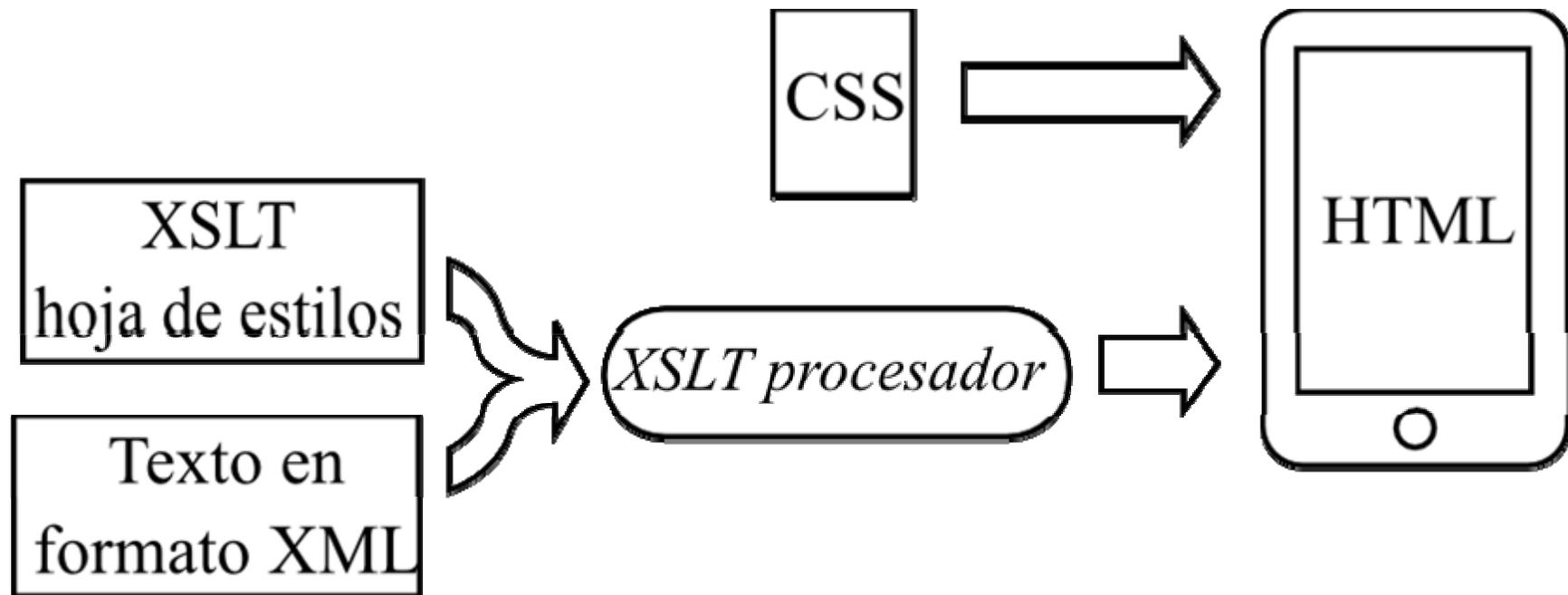
y no en:

- Los requisitos técnicos de un sistema diseñado para otros objetivos
- La estructura de datos requerida por programas generales de formateado

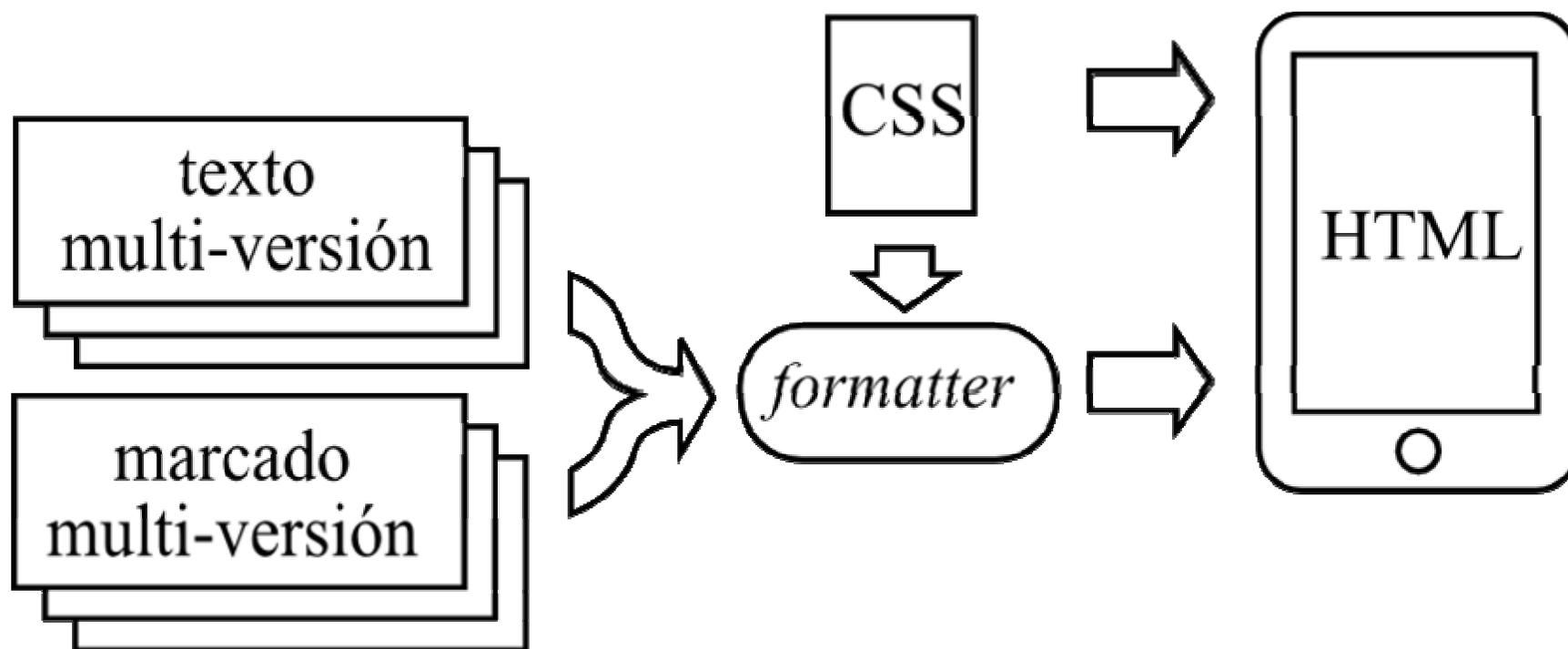
HRIT vs XML

- Las HD se caracterizan por el uso de XML y de las herramientas XML (XSLT, XQuery, Oxygen etc.).
- Tal vez TEI-XML se haya convertido en una tecnología “lock-in”?
- HRIT evita el uso de XML para la representación de los datos fundamentales.
- Cómo lo hace?

Infraestructura digital con XML



Infraestructura HRIT



Principales ventajas de HRIT

- El marcado está separado del texto
- Texto y marcado pueden existir en múltiples versiones
- Las unidades marcadas pueden combinarse libremente y ser reutilizadas para otras finalidades
- Las unidades de marcado se pueden formatear en HTML para la visualización. La conversión está controlada por un fichero CSS estándar (NO XSLT!)

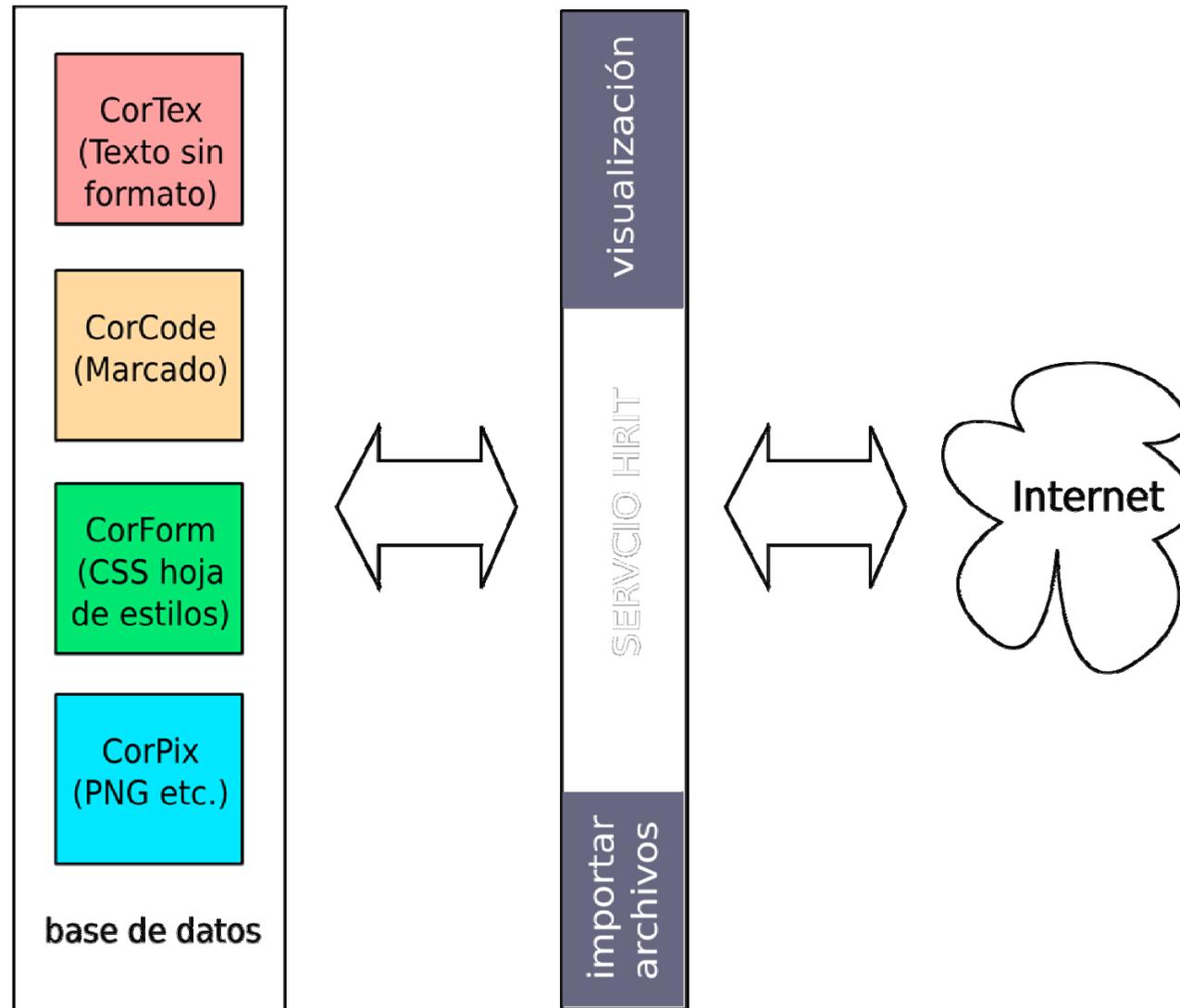
Ventajas 2

- El marcado puede solaparse (*overlap*) libremente porque no está vinculado a una estructura jerárquica.
 - Las jerarquías solapantes (*overlapping hierarchies*) están consideradas uno de los problemas que no encuentran una solución satisfactoria en XML.

Ventajas 3

- La separación entre texto y marcado aumenta la usabilidad.
 - Y editar texto simple es más fácil...
- Ya no necesitamos marcar manualmente las variantes.
- Las versiones se gestionan automáticamente.
- La visualización de la variación es más efectiva (ej. side by side view, table view).

Visión general del diseño de HRIT



Interoperabilidad y usabilidad

- HRIT solo requiere que el usuario o el investigador lea o comprenda HTML.
- Podemos usar cualquier tipo de herramienta y de lenguaje de programación.
- Imágenes y texto plano son altamente interoperables y reutilizables.
- La interoperabilidad del mercado se ve incrementada gracias a la posibilidad de nuevas utilizaciones.

Proyectos en marcha

- Digital Variants (2003-) – creación del concepto de Multi-Version Document.
- [HRIT](#) en Loyola University, Chicago (2009-) – basado en ediciones impresas, añadido el modelo *standoff markup*.
- AustESE (electronic scholarly editions) University of Queensland Australia (2011-) – colección de textos impresos y manuscritos, financiado a 3 años con 3 programadores.
- Rabindranath Tagore electronic edition (2011-) Jadavpur University, Kolkata, India: obras de teatro, novelas, música.