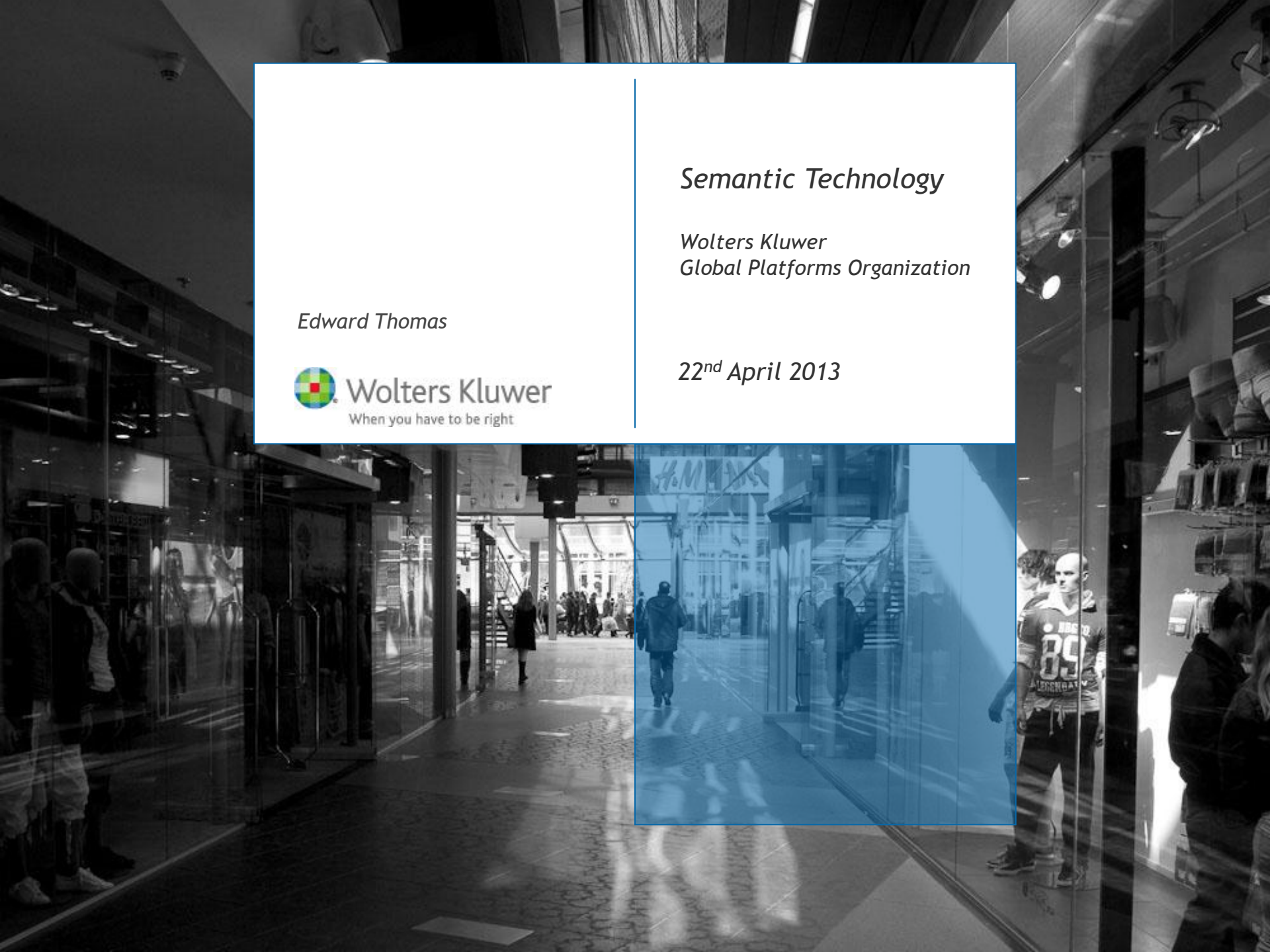


Semantic Technology

Wolters Kluwer
Global Platforms Organization

Edward Thomas

Wolters Kluwer
When you have to be right

22nd April 2013

Presenter

Edward Thomas
Applied Research Engineer
Edward.Thomas@wolterskluwer.com

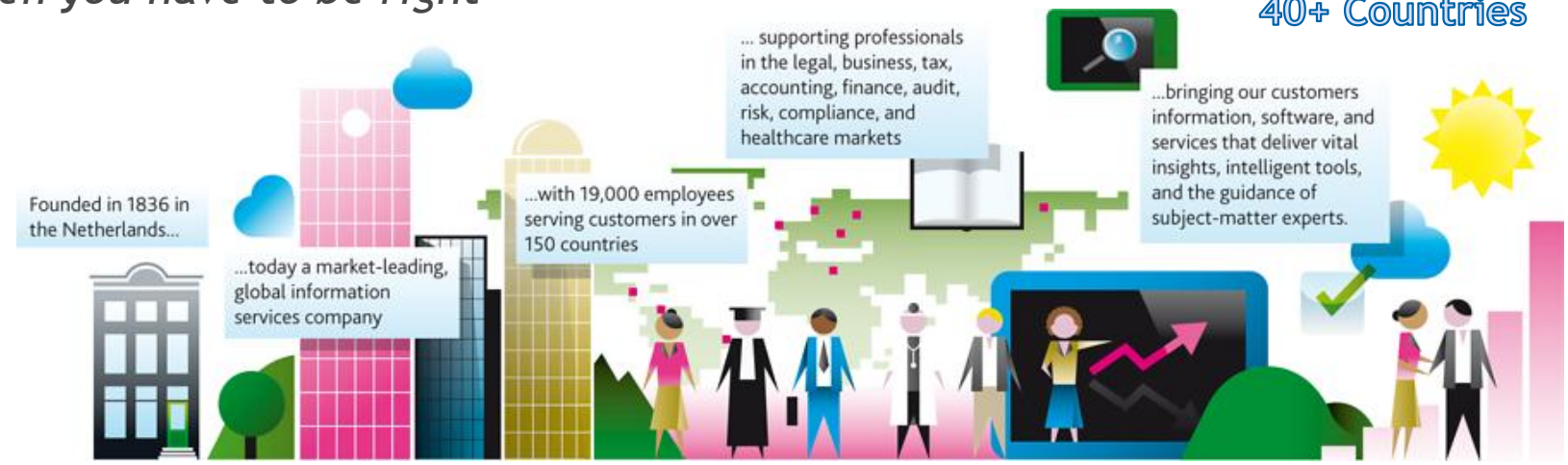
Wolters Kluwer
Global Platforms Organization

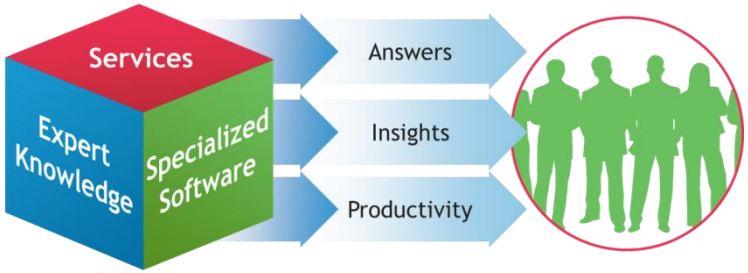
As part of the Advanced Technologies Group in GPO, I am responsible for finding promising new technologies and applying them to business problems at Wolters Kluwer.

Wolters Kluwer

When you have to be right

40+ Countries





Wolters Kluwer supports **90% of U.S. banks** and the top **40 global banks**

Wolters Kluwer supports over **210,000 tax & accounting firms worldwide** that's over 40 million tax returns

Wolters Kluwer 2012	
Revenue	€ 3.6 billion
Ordinary EBITA	€ 785 Mln
Employees	18,400
Free cash flow	€ 510 Mln

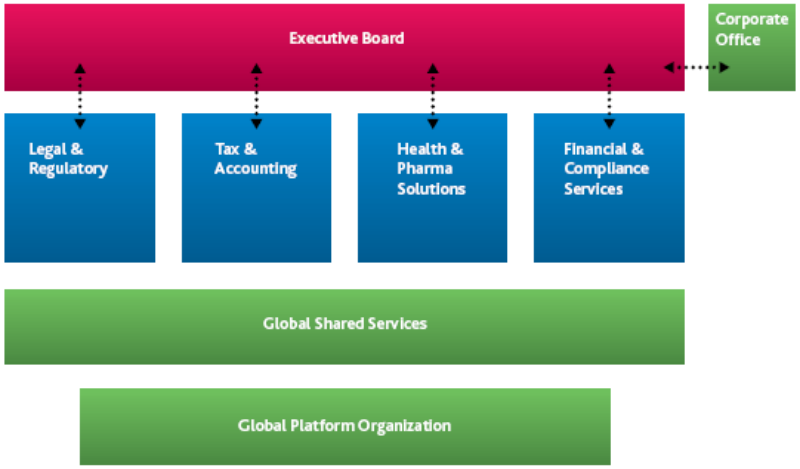
Wolters Kluwer supports over **13 million healthcare professionals** in more than 150 countries

Wolters Kluwer supports over **250,000 legal professionals** worldwide

Markets: Legal, Business, Tax, Accounting, Finance, Audit, Risk, Compliance, and Healthcare

Wolters Kluwer: Global Platforms Organization

- Managed by: Dennis Cahill, Executive Vice President
- 108 Employees
- Supports innovative solutions development across the company



- Global Solutions/Services around:
  - Search, Mobile, Current Awareness, Content & Knowledge Management, Syndication, Software & Knowledge Integration, Content Federation, UX Design, etc.
- Responsible for supporting the Wolters Kluwer business with their innovation, while driving for generalized solutions around the globe

What does Wolters Kluwer expect from *Semantic Web Technologies*

- Better products for our customers
  - Better integration with non-Wolters Kluwer data
  - More focused solutions to actual information needs
  - Better integration into the customer's workflow
- Improved nimbleness in following market demands
  - Easier syndication of data and knowledge repositories
  - Simplified integration of data and software
  - Improved capabilities in mapping internal and external vocabularies
- Increased operational efficiencies
  - By not having to do what has been done elsewhere
  - By applying the concept of economies of scale even in a semantic diverse context

Key Characteristics of our Content Model

The architecture is…

- Semantic
- Metadata centric
- Standards based (OWL, RDF, XHTML+RDFa, SKOS)
- Natively extendable
- Uses ontologies

What is semantic about our model?

- Core model covers publishing concepts
  - What is a document?
  - What is a citation?
  - etc.
- Additional domain models cover e.g.
  - Legal publishing artifacts
  - Legal concepts
  - ...

The Content Architecture Supports

- The Wolters Kluwer organisation
  - Core Model + Business specific extensions
  - 4 divisions in 150 countries
- Diverse content sources
  - Editorial content
  - Open data
  - External content
  - User contributed content
  - Customer data
- Diverse channels
  - Web & Mobile
  - Workflow and decision support
  - Compliance Reporting support
  - ePub/Book/Loose-leaf

Content Specification

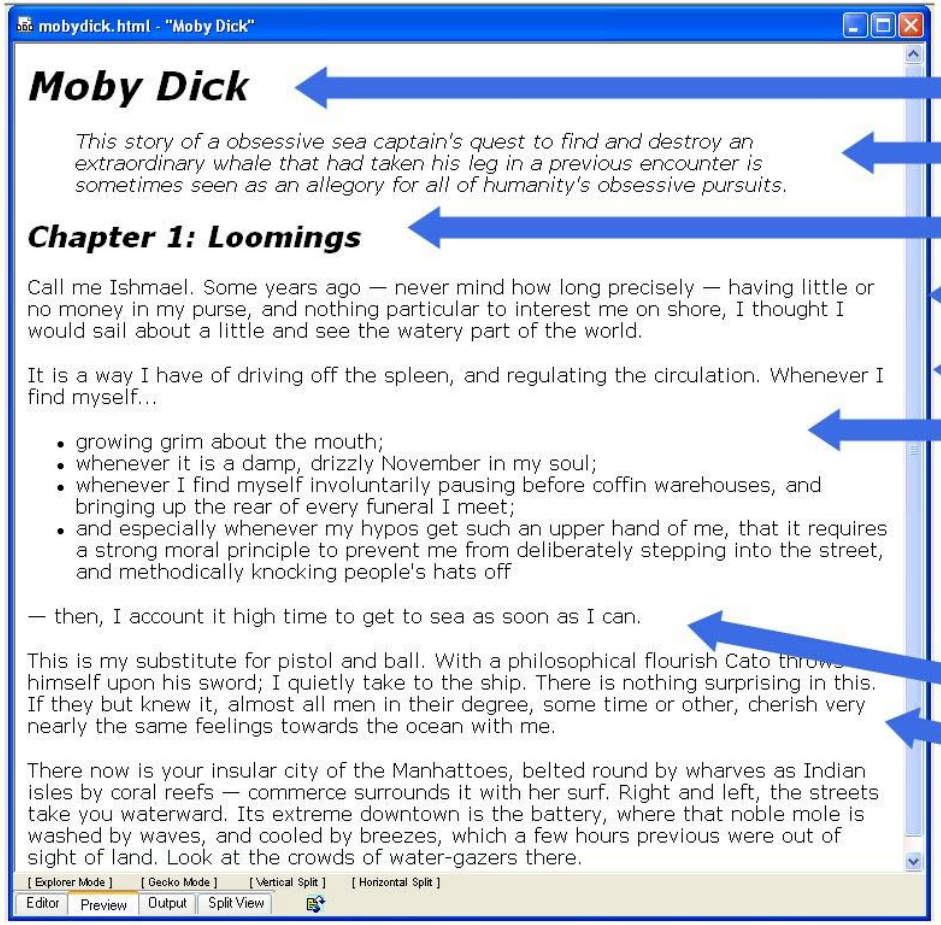
- Encodes textual component of documents
- Based on W3C XHTML 1.1 specification, which is
  - Conformed to the XML standard
  - Designed to separate content from presentation
  - Extensible enabling new elements to be defined
- Our Content Specification differs from XHTML by
  - Removal of modules not required for WK content (e.g. forms, events)
  - Addition support for RDF annotations (RDFa)

Content Specification - RDFa

- Uses XHMTL tags to express RDF triples
- Provides syntax (@resource, @about, @typeof, etc.) for expressing RDF triples in XHTML

```
<div xmlns:v="http://rdf.data-vocabulary.org/#"
typeof="v:Person">
<span typeof="v:Address">
<span property="v:locality">Albuquerque</span>
<span property="v:region">NM</span>
</span>
</div>
```

Content Specification – Content Structure



Content Structure

- Hierarchy of document content (eg. US Code)
  - Paragraph partOf SubSection
  - SubSection partOf Section
  - Section partOf Part
  - Part partOf SubChapter
  - SubChapter partOf Chapter
  - Chapter partOf SubTitle
  - SubTitle partOf Title
- All can have headers, abstracts, annotations, citations, relationships, etc.

Benchmark use cases at Wolters Kluwer

- Content Management platform
- Publishing platform

Content management

- Read write context
- Editorial operations, content enrichment operations,
- Good mix of CRUD
- Heavily transactional
- Heavy use of OWL/RDFS
- Validation of data in place using ontologies/schemas
- Data is packaged and sent to the publishing system based on a complex content specification

Publishing

- Overwhelmingly read-only
  - Read queries on access
  - Scheduled writes of large volumes of data from CMS
- Faceted full-text search based on metadata + lightweight reasoning
- Large scale
  - 16M documents
  - ~1000 triples per document
  - 10 000s of daily users
- Fine grained entitlement checking

General requirements

- Availability
  - 99.99% availability for publishing platform
  - 99% availability for content management platform
- Scalability
  - Number of users is stable
  - Quantity and complexity of data is growing
  - Number of applications of data is growing

What else?

- Internal benchmarking is important to us
  - Lots and lots of proprietary content
  - This includes the model, makes it difficult to anonymize
- Access to tools to support this would be very helpful

More information or questions?

Edward Thomas

<edward.thomas@wolterskluwer.com>