New Directions in Content Modeling

Learning from Open World Content

Marc Salvatierra

Open Source CMS 26 conference
20 October 2026 | Utrecht, Netherlands

Marc Salvatierra

- **Content –** news, media, entertainment, business, legal, tech
- **Where –** CH1, Sony, Cox, eToys, Viacom, Auto Club, O'Melveny
- **ICANN: 2006 → now –** 30+ sites; many content and Web roles

Channel One News

SONY ONLINE ENTERTAINMENT

eToys.com

Automobile Club of Southern California AAA

ICANN

O'Melveny

COX INTERACTIVE media

A Subsidiary of Cox Enterprises, Inc.

VIACOM

Marc Salvatierra

CMS DMS Product Semantics

Hummingbird DOCS Open

LAInsider PacMan

OMM Registration Grafo dotCMS Tektronix NewStar

Rails apps Zope Notepad++

WorldNow Alfresco

Drupal

Movable Type WEBVowl

Relativity

protégé sFTP

Wordpress

Squarespace

Textpattern

Introduction
Goals

Examine “open world” content
beyond the CMS content model.

Appreciate content modeling history
from ancient writing to digital practices.

Explore possible new directions for
content modeling and content systems.

Marc Salvatierra

Introduction
Thesis

Open world content around
us endures and often dominates.

**The CMS content model faces
pressure to grow in scope and specificity.**

**Adjacent disciplines offer possibilities for
future content modeling and content systems.**

Open World Content

Remediation … Remodeling

"the content of any medium is always another medium"

– Marshall McLuhan, 1964

Remediation: "the formal logic by which new media refashion prior media forms"

– Jay David Bolter & Richard Grusin, 1999

The Digital Age: "new media transform older media, retaining some of their features while discarding others"

– Philipp Schweighauser, 2023

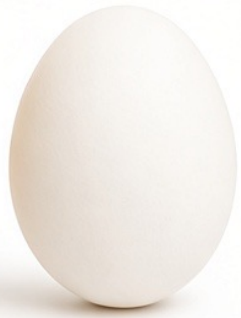
Marc Salvatierra

Open World Content
A Definition

"Open World" content is modeled from content that already exists.

"Closed World" content is created within a model that already exists.





Marc Salvatierra

Open World Content
In the Wild

content outside *our* CMS content model and channels;
content that is macro or micro to the page dimension

non-digital, historical, literary and linguistic content;
corpora in the digital humanities and digital scholarship

ephemeral information objects; organic aggregations;
bibliographic artifacts; environmental installations & DOOH

Marc Salvatierra

Open World Content Aspects

Making Before Modeling – Exists before its essence examined.

Presentation As Content – Conveyed through multimodality.

Abundance, Bustle, Collage – Unstructured in its aggregation.

Self-Description – Includes peritext from author and publisher.

Groundwork Coffee – Venice, California

Marc Salvatierra

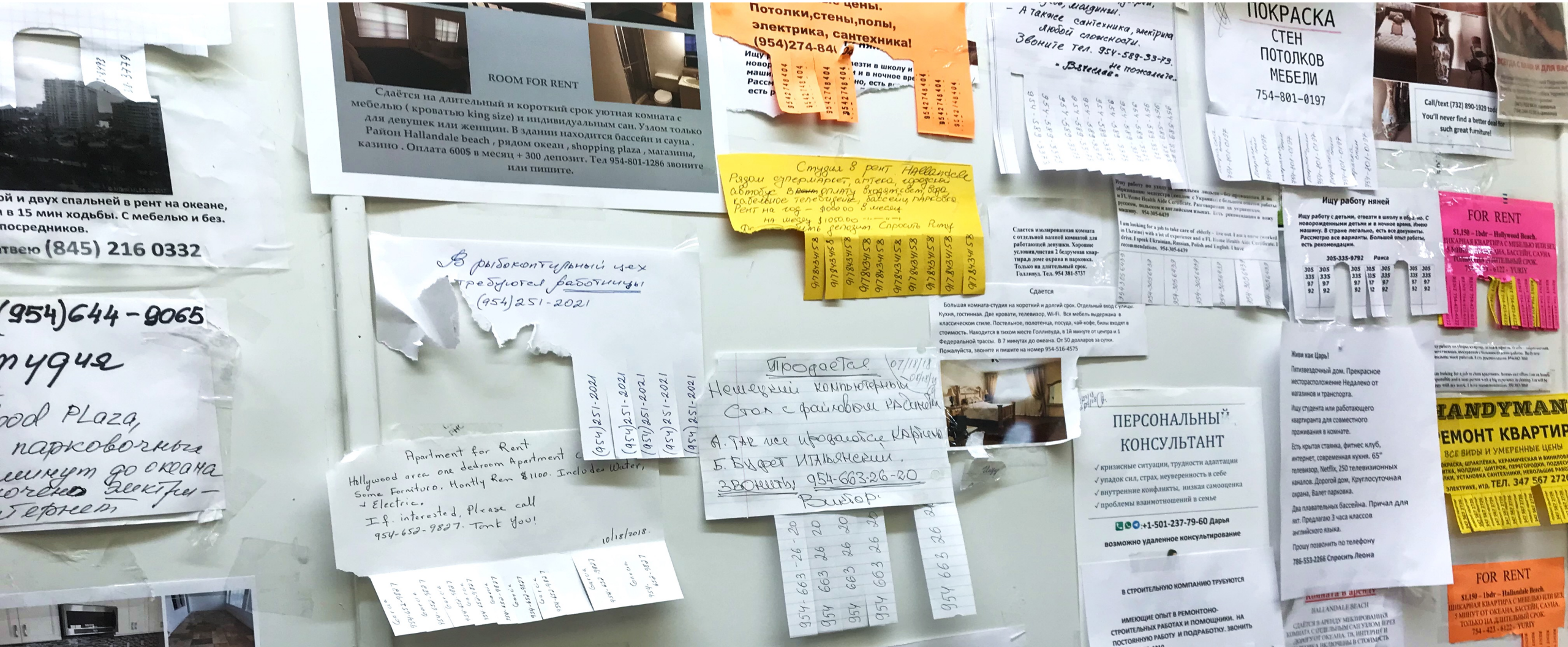
Open World Content
Ephemeral Objects



World Market & Café – Hollywood, Florida

Open World Content
Ephemeral Objects

Marc Salvatierra



Ink@84 Books – Highbury Park, London

Marc Salvatierra

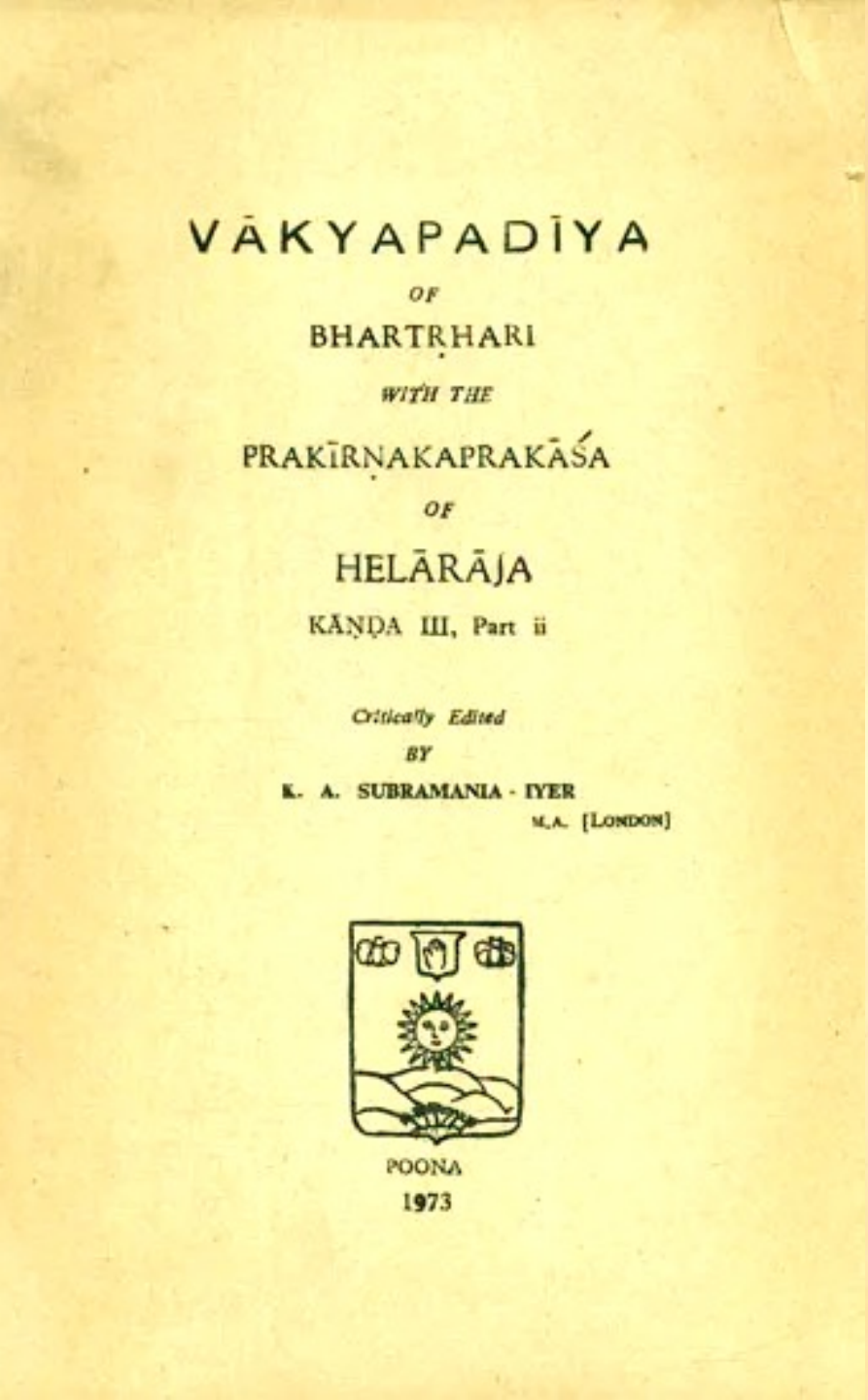
Open World Content
Ephemeral Objects



Marc Salvatierra

Open World Content

Bibliographic Artifacts


VĀKYAPADĪYA

OF

BHARTṚHARI

WITH THE

PRAKĪRṆAKAPRAKĀŚA

OF

HELĀRĀJA

KĀṆḌA III, Part ii

Critically Edited

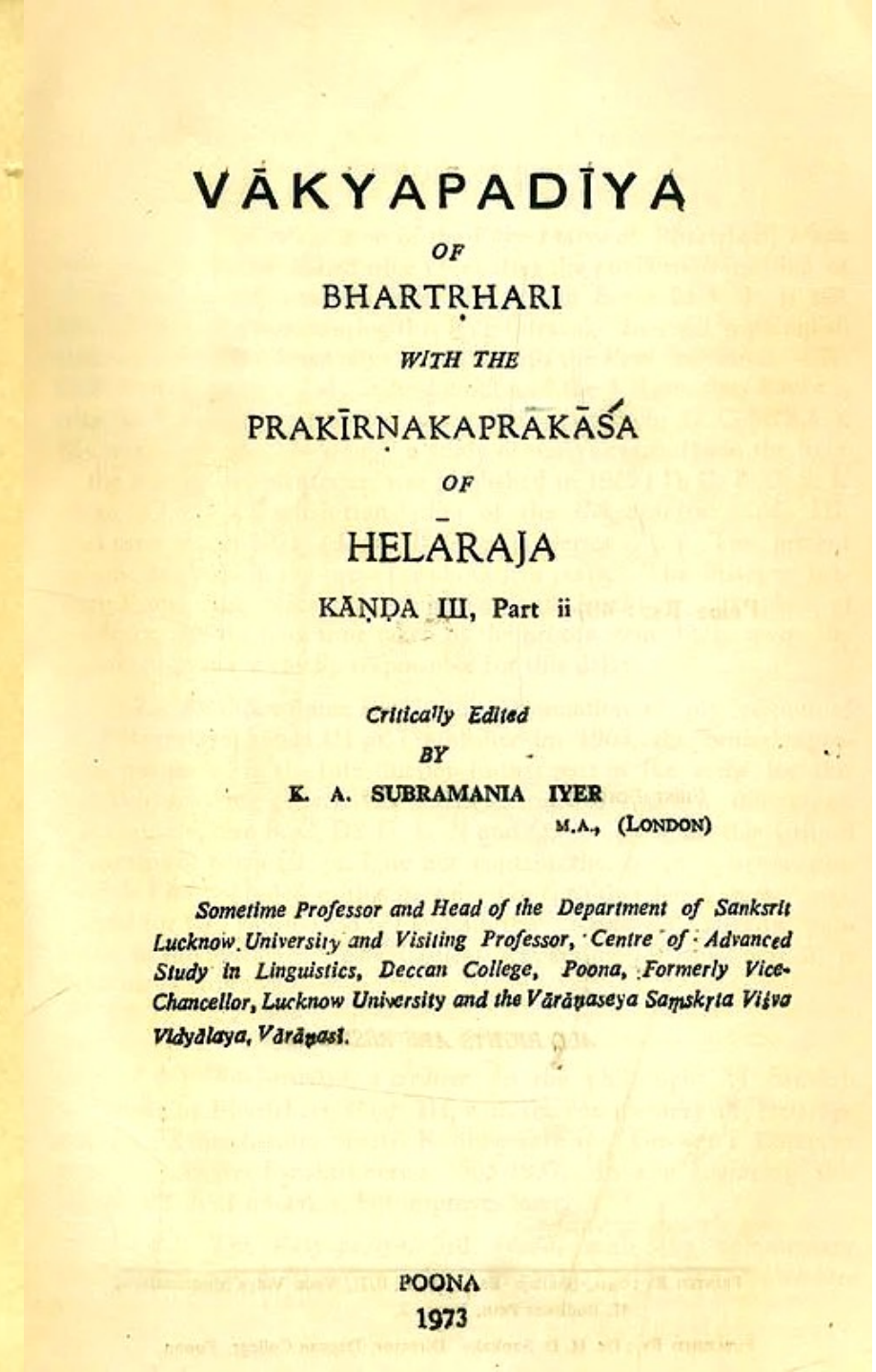
BY

K. A. SUBRAMANIA - IYER

M.A. [LONDON]

POONA

1973


VĀKYAPADĪYA

OF

BHARTṚHARI

WITH THE

PRAKĪRṆAKAPRĀKĀŚA

OF

HELĀRAJA

KĀṆḌA III, Part ii

Critically Edited

BY

K. A. SUBRAMANIA IYER

M.A., (LONDON)

Sometime Professor and Head of the Department of Sanksrit Lucknow University and Visiting Professor, Centre of Advanced Study in Linguistics, Deccan College, Poona, Formerly Vice-Chancellor, Lucknow University and the Vārāṇaseya Saṃskṛta Viśva Vidyālaya, Vārāṇasi.

POONA

1973

Vākyapadīya
Kāṇḍa III, Part ii
Bhartrhari
5th century CE

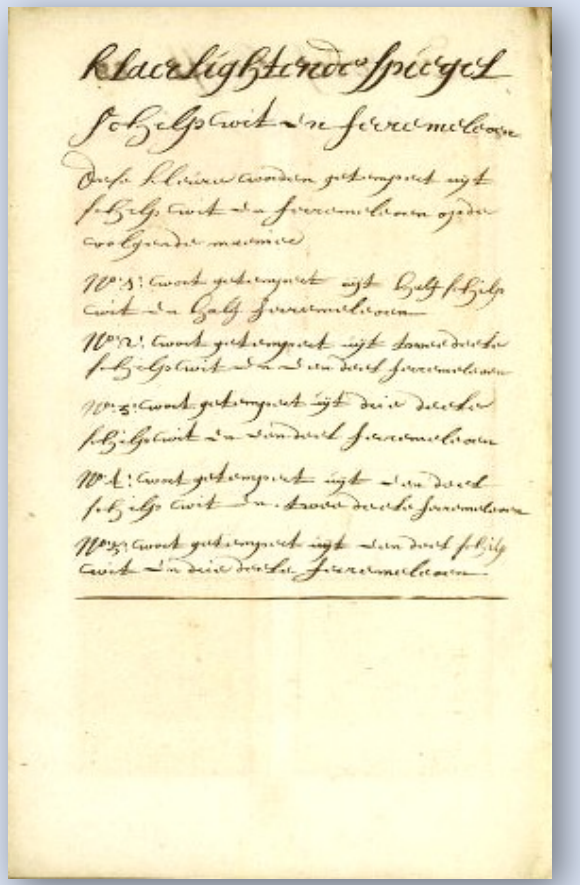
critical edition, K. A. Subramania Iyer, Poona, 1973

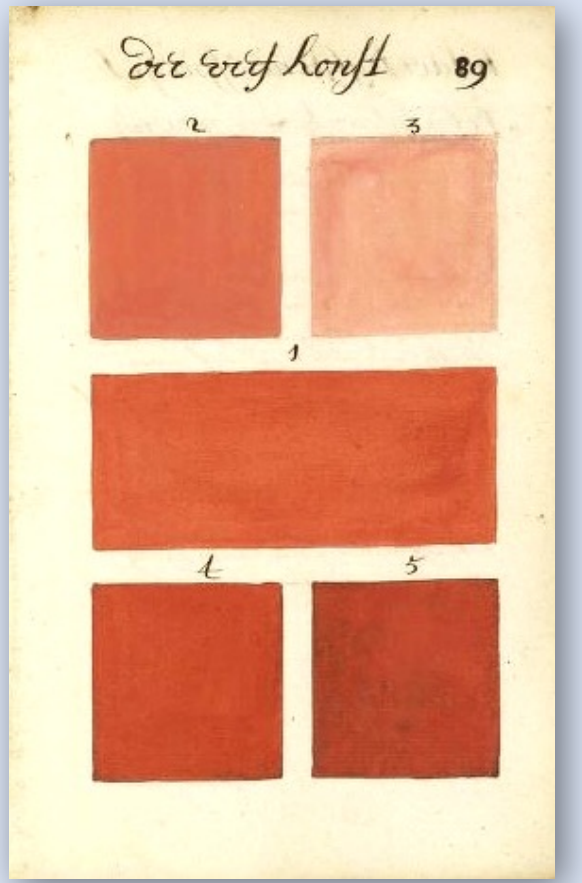
Exotic India Art Pvt Ltd

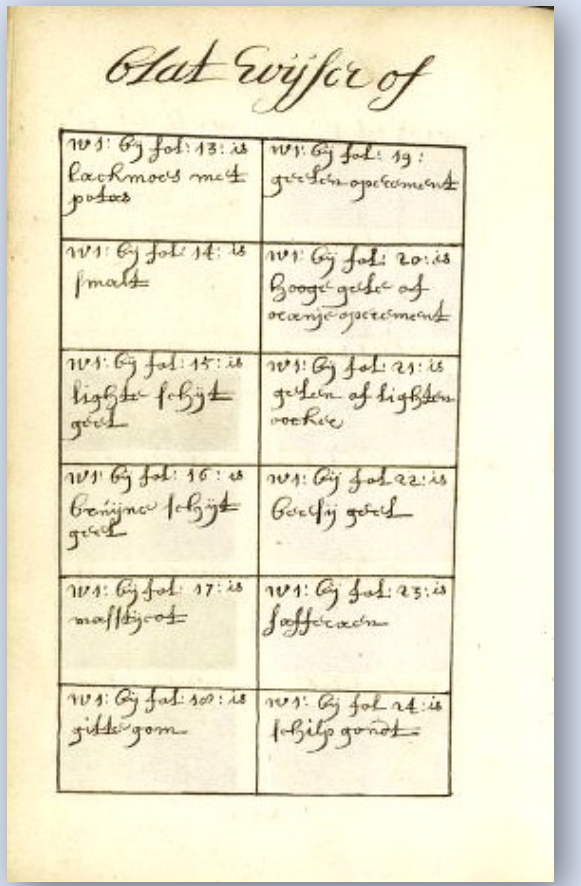
Marc Salvatierra

Traité des couleurs – A. Boogert
1692

Open World Content
Bibliographic Artifacts









Marc Salvatierra

Open World Content
Bibliographic Artifacts

The Studio Zine – Teesha Moore
1990s – 2000s

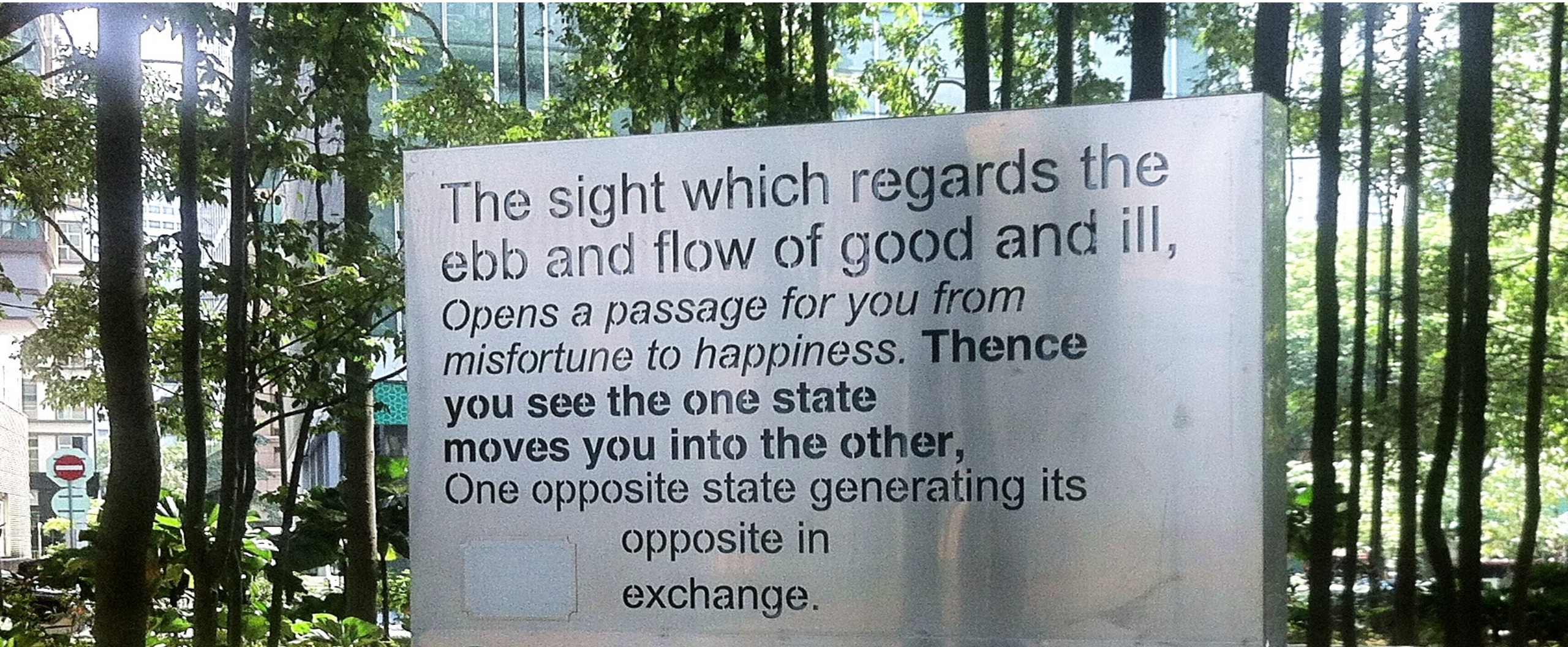
Stone-Hausmann, eBay



Kuala Lumpur, Malaysia

Marc Salvatierra

Open World Content
Environmental Installations



Times Square – New York City

Open World Content
Environmental Installations

Marc Salvatierra



Sunset Boulevard – Hollywood, California

Open World Content
Environmental Installations

Marc Salvatierra



Toward Content Types
Theory & Formation

Stock of Knowledge – A person interprets life through an adaptable set of socially acquired ‘typifications’. (Alfred Schütz)

Genre as Social Action – Human behavior normalizes and solidifies in response to ‘recurring situations’. (Carolyn Miller)

Document Theory – An object assumes document status when treated as evidence of something; documents materialize information. (Suzanne Briet, Michael Buckland, Paul Otlet)

Marc Salvatierra

Toward Content Types

htp-dj-nsw 



"An offering that the king gives”

standard dedicatory text found primarily on funerary items

Egypt | c. 2600 BCE (Old Kingdom) to Ptolemaic / Roman period

format :

1. King’s Formula – Royal Gift
2. God’s Formula – Divine Invocation
3. visitor’s voice invocation formula
4. inventory of items offered
5. name & title of beneficiary

“Ancient Egyptian Offering Formula,” Wikipedia, https://en.wikipedia.org/wiki/Ancient_Egyptian_offering_formula

Marc Salvatierra

Toward Content Types

ACTA·DIVRNA



iStock

official daily Roman gazette

chronology of government acts and announcements, court proceedings, elite family occurrences, gladiatorial news

Rome | circa 59 BCE – 222 CE

publication :

1. posted recurringly, publicly
2. likely prepared on a whitewashed surface
3. unlikely metal or stone
4. hand-copied
5. no surviving copies

Marc Salvatierra

Toward Content Types

letters of indulgence

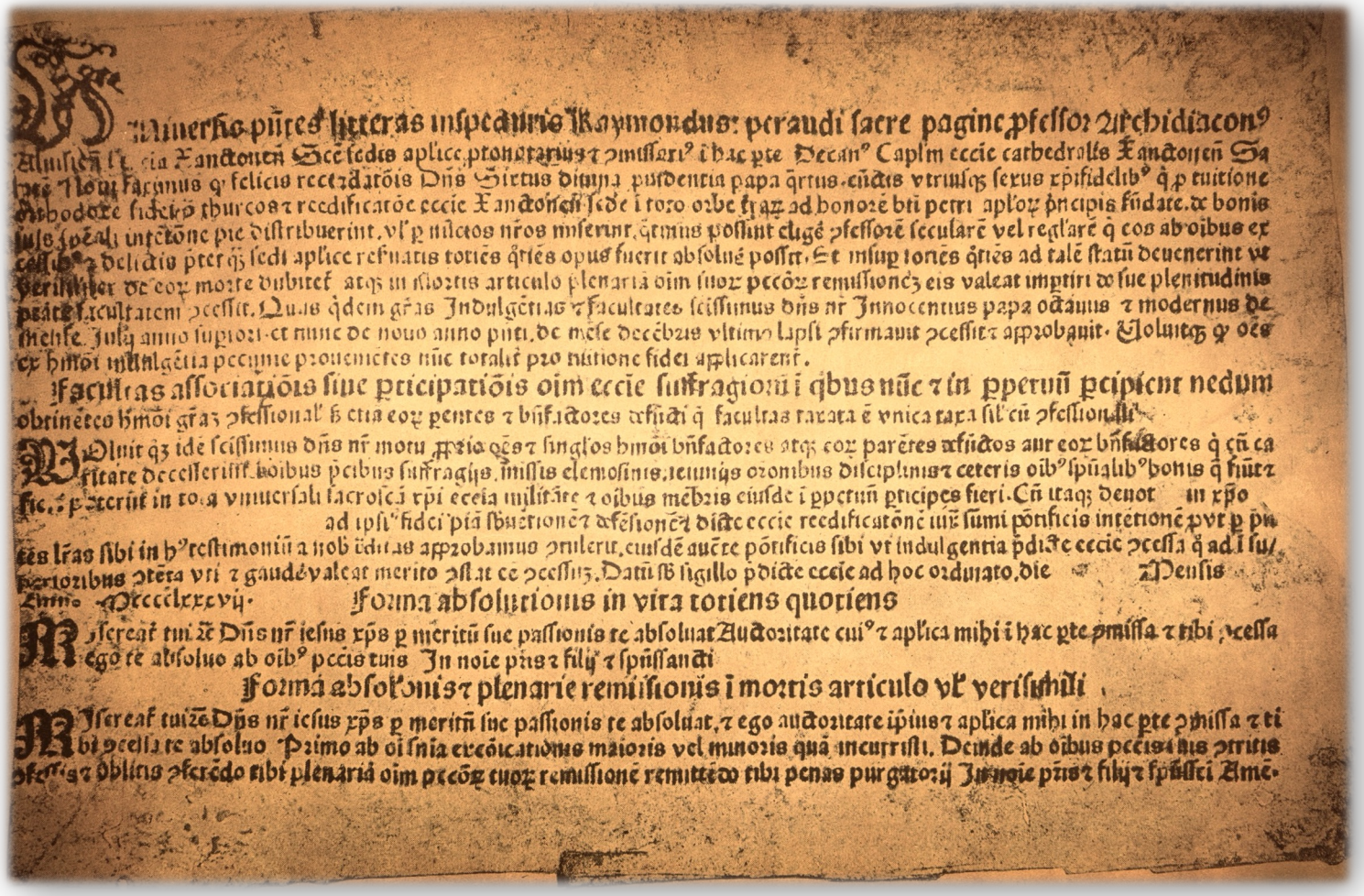
Catholic certificate of an indulgence given to remit temporal punishment for sin

structured document written in Latin on parchment, later printed

Europe | circa 1300 – 1570 CE |
Late Middle Ages to Reformation

fields used :

1. invocation
2. issuer name & authority
3. recipient | beneficiary name(s)
4. greeting
5. preamble & motivation
6. the grant
7. conditions
8. absolution formula(s)
9. date, place
10. signature & seal



Forma absolutionis in vita toriens quoriens

Forma absolutionis et plenarie remissionis in mortis articulo vt versutibili

blank letter of indulgence issued under Pope Innocent VIII, circa 1487 – 1488 CE

iStock

Toward Content Types

dibao 邸報



reports from provincial liaison residences in the capital; later controlled by the central court

court gazettes – one of the world's earliest quasi-newspapers

China | circa 8th century – early 20th century CE, Tang – Qing dynasties; names varied throughout

publication :

1. handwritten, later also printed, compiled from court documents
2. circulated to officials across China, later to wider readers
3. Qing format: emperor's activities, audiences, travel, appointments; imperial edicts, memorials

Chinese Rare Book Collection. *Jing Chao: Si Shi Er Ce*. [China: s.n., Qing Daoguang 22 zhi 23 nian, 1842 – 1843] Manuscript / Mixed Material. https://www.loc.gov/item/2012402827/

Toward Content Types

IBM General Document

```
Overall Structure:
    General Document (GDOC)
        front matter (FRONTM)
            title page (TITLEP)
                document title (TITLE)
                document number (DOCNUM)
                document date (DATE)
                author name (AUTHOR)
                address of author or publisher (ADDRESS)
                    address line (ALINE)
            abstract (ABSTRACT)
            preface (PREFACE)
            table of contents (TOC)
            list of illustrations (FIGLIST)
        body (BODY): May contain either parts or chapters
            part (implied by H0)
                head zero (H0)
                basic document elements
                chapter (implied by H1)
                    head one (H1)
                    basic document elements
                    lower-level heading segments (implied by H2 to H6)
        appendix section (APPENDIX)
            first-level heading segment (implied by H1)
        back matter (BACKM)
            first-level heading segment (implied by H1)
```

default document type included in IBM's Document Composition Facility

USA | released by IBM in 1978

features :

1. basic structure for text files destined for print output
2. shipped with a default, expandable tag vocabulary
3. supported wide range of logical structures
4. commonly adopted within the IBM customer base

International Business Machines Corporation, *Document Composition Facility: Generalized Markup Language (GML) User's Guide*, 1st ed., publication no. SH20-9160-0 (IBM, 1978), page 7

Toward Content Modeling

"Intentional Arrangement" (Glushko)

"explicit or implicit acts of organization by people, or by computational processes acting as proxies for, or as implementations of, human intentionality"

– Robert Glushko

Toward Content Modeling

Why Do We Model Content?

Telling & Transacting – Spectrum of prose for reading and structured data for processing. (Glushko & McGrath, *Document Engineering*)

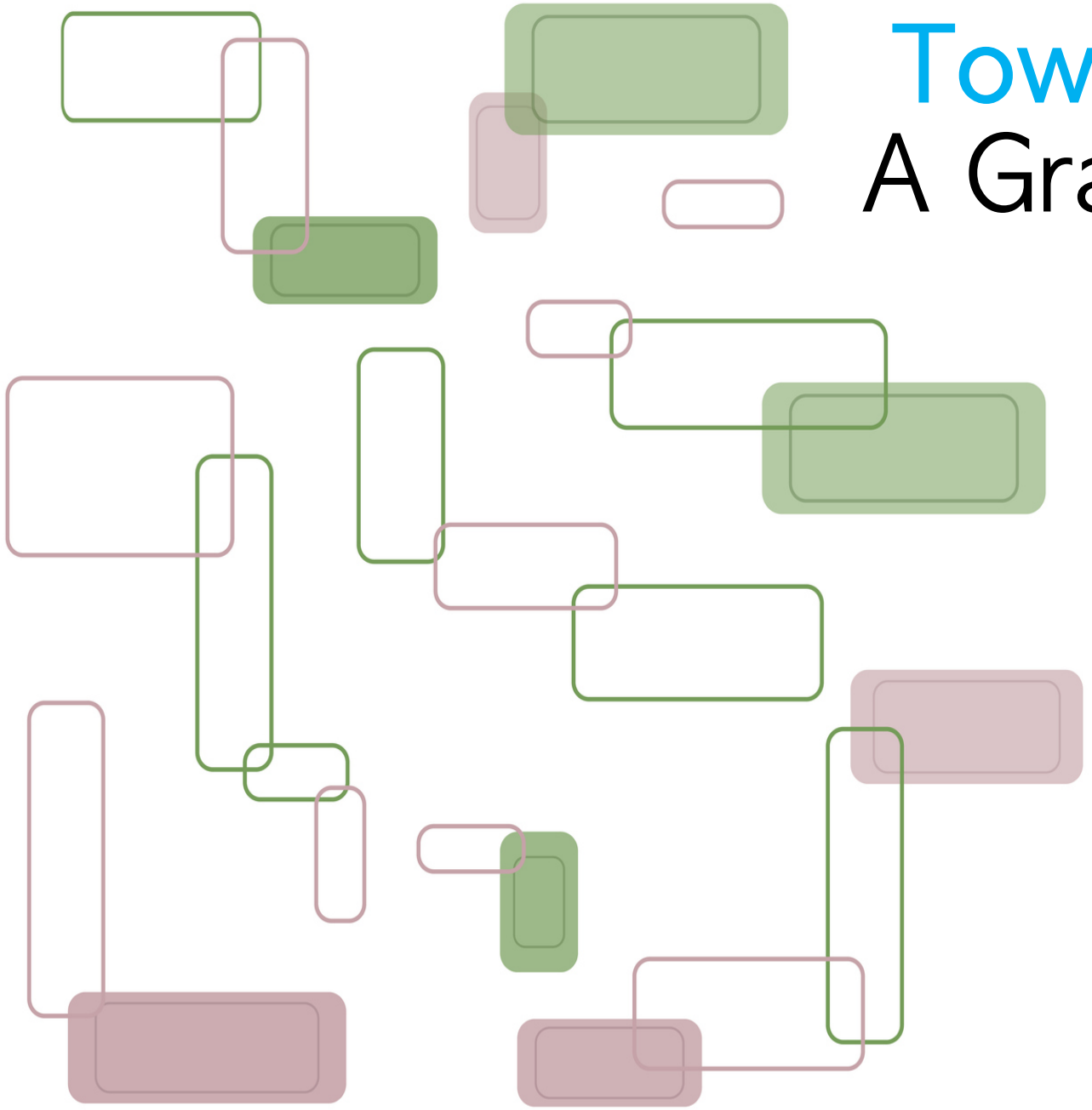
Principles – "digital environments … reflect our values, our beliefs and our aspirations" (Talisman, "Intentional Arrangement – A Manifesto")

Capabilities – "To *organize* is to create capabilities by intentionally imposing order and structure." (Glushko, *The Discipline of Organizing*)

Marc Salvatierra

Toward Content Modeling

A Grammar of Documents



"There appears to be a set of structural features common to the majority of text documents that have become a part of the way the human race has recorded textual information over the millen[n]ia."

– Peter Flynn

Table I

(Non-Periodic) "Table of the Elements" from selected XML grammars (LaTeX has been included for comparison)

Feature	HTML	DocBook	DITA	TEI	12083	JATS	Briefing	Bulletin	LaTeX
title	title	title	title	title	title	article-title	title	title	\title
author		author	author	author	author		briefeditors	author	\author
summary		abstract	shortdesc	abstract	abstract	abstract		abstract	abstract
preface		preface		front	preface				\frontmatter
part		part	section	div|div0	part	sec			\part
chapter	h1	chapter	section	div|div1	chapter	sec		report	\chapter
section	h2	sect1	section	div|div2	section	sec	story	section	\section
subsection	h3	sect2	section	div|div3	subsect1		level3	sub.section	\subsection
subsubsection	h4	sect3	section	div|div4	subsect2		level4		\subsubsection
appendix		appendix		appendix	afterwrd				\appendix
bibliography		bibliography		listBibl	biblist	Ref-list	biblist		thebibliography
index		index		index	index				index
glossary		glossary		glossary	glossary	glossary	glosslist		glossary
paragraph	p	para	p	p	p	p	para	ptxt	\par
quotation	blockquote	blockquote	lq	quote	bq			block.quote	quotation
numbered list	ol	orderedlist	ol	list	list	list	numberlist	list	enumerate
bulleted list	ul	itemizedlist	ul	list	list	list	bulletlist		itemize
dictionary list	dl	variablelist	dl	list	deflist	list	defnlist		description
figure	img	figure	fig	figure	fig	fig	illus	figure	figure
table	table	table	table	table	table	table	table	table	table
mathematics		equation		formula	formula	mml:math	formula		$$
cross-reference	a	xref	link	ref	secref		xref	eiro.ref	\ref
bibliographic reference	a	biblioref	cite	ref	citeref			biblio	\cite
external link	a	link	xref	ptr			weblink	external.ref	\hyperref
emphasis	em	emphasis		emph	emph		emph	emph1	\emph
language	lang	foreignphrase		foreign				language.phrase	\selectlanguage

Peter Flynn, "Your Standard Average Document Grammar: Just Not Your Average Standard," in *Proceedings of Balisage: The Markup Conference 2017*, Balisage Series on Markup Technologies, vol. 19 (2017)

A Grammar of Documents

"Table of the Elements"

Peter Flynn

A Grammar of Documents

IBM General Document

Type	Attributes	Content	Termination
:abstract	None	No Immediate Text	:preface, :toc, :figlist, or :body
:address	None	No Immediate Text	:eaddress
:aline	None	Text on Same Line (no other tags)	At End of Line
:appendix	None	No Immediate Text	:backm or :egdoc
:author	None	Text on Same Line (no other tags)	At End of Line
:backm	None	No Immediate Text	:egdoc
:body	None	No Immediate Text	:appendix, :backm, or :egdoc
:date	None	Text on Same Line (no other tags)	At End of Line
:dd	None	Implied P Structure	:dt, :lp, or :edl
:dl	termhi= tsize=	No Immediate Text	:edl
:docnum	None	Text on Same Line (no other tags)	At End of Line

International Business Machines Corporation, *Document Composition Facility: Generalized Markup Language (GML) User's Guide*, 1st ed., publication no. SH20-9160-0 (IBM, 1978), pages 134–135

Toward Content Modeling Diplomatics

Establishing Authenticity – Method created by Jean Mabillon in *De re diplomatica* in 1681 to evaluate disputed charters.

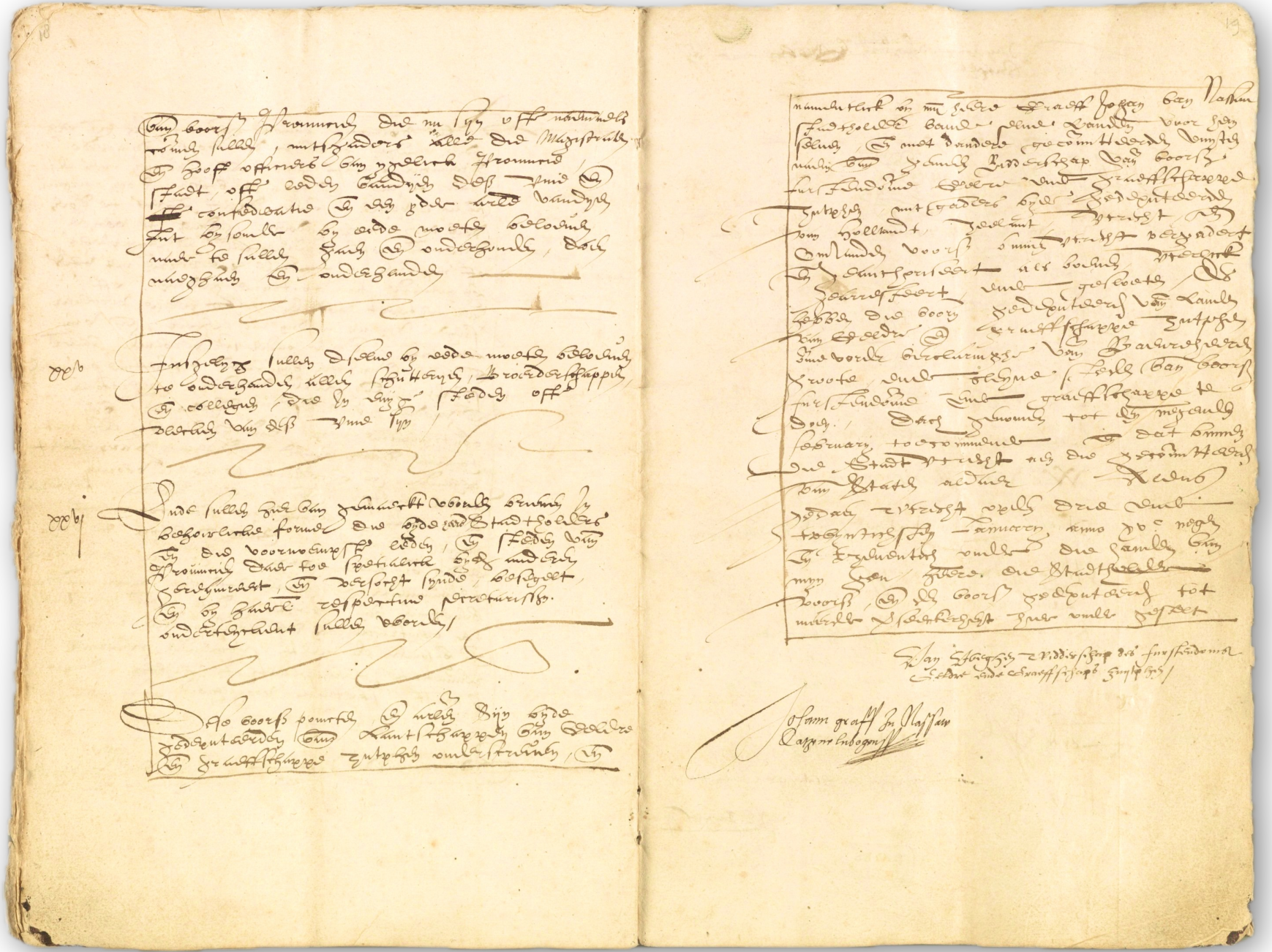
Structural Analysis – Later diplomatists formalized the division of the charter into protocol, text and eschatocol.

Standard Clauses – invocatio, intitulatio, inscriptio; arenga, narratio, dispositio, sanctio, corroboratio; datum

Marc Salvatierra

Diplomatics

Union of Utrecht – excerpt



Nationaal Archief,
Nederland

circa 1579 – 1580

Toward Content Modeling

Library & Archival Practices

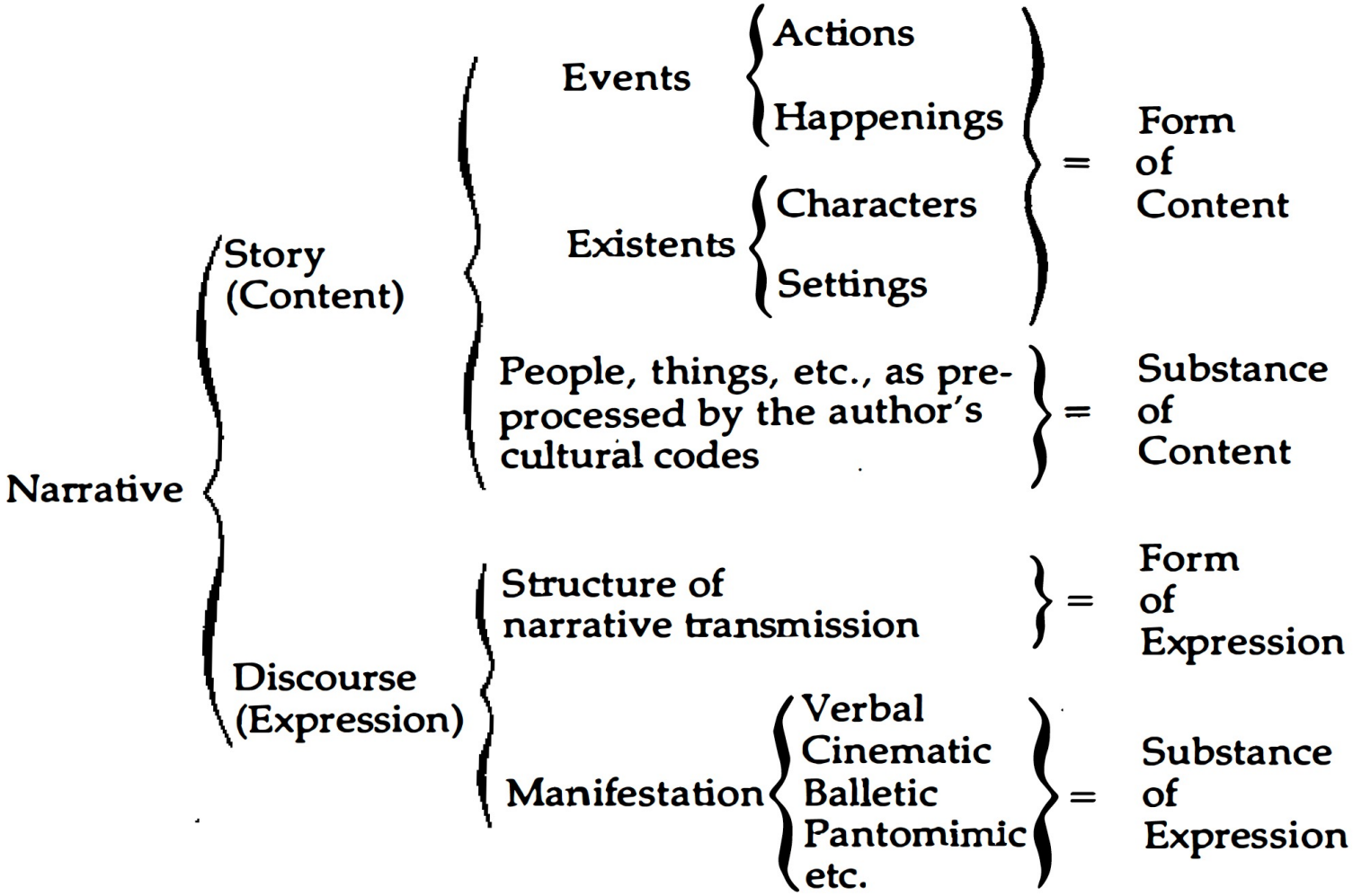
Respect des Fonds – Keep a creator's records separate from other creators' records. (Natalis de Wailly, 1841)

Original Order – Preserve records in the arrangement inherited from their creator (Max Lehmann, 1881).

Monographic Principle – Decompose documents into discrete facts; record them separately for reassembly. (Paul Otlet, 1918)

Toward Content Modeling

Classical Narratology



1960s: Seeks the universal "grammar" of narrative – structures common to all stories – regardless of origin, content or society.

diagram of narrative, *Story and Discourse – Narrative Structure in Fiction and Film* – Seymour Chatman, 1978

Marc Salvatierra

Toward Content Modeling

A Grammar of Models

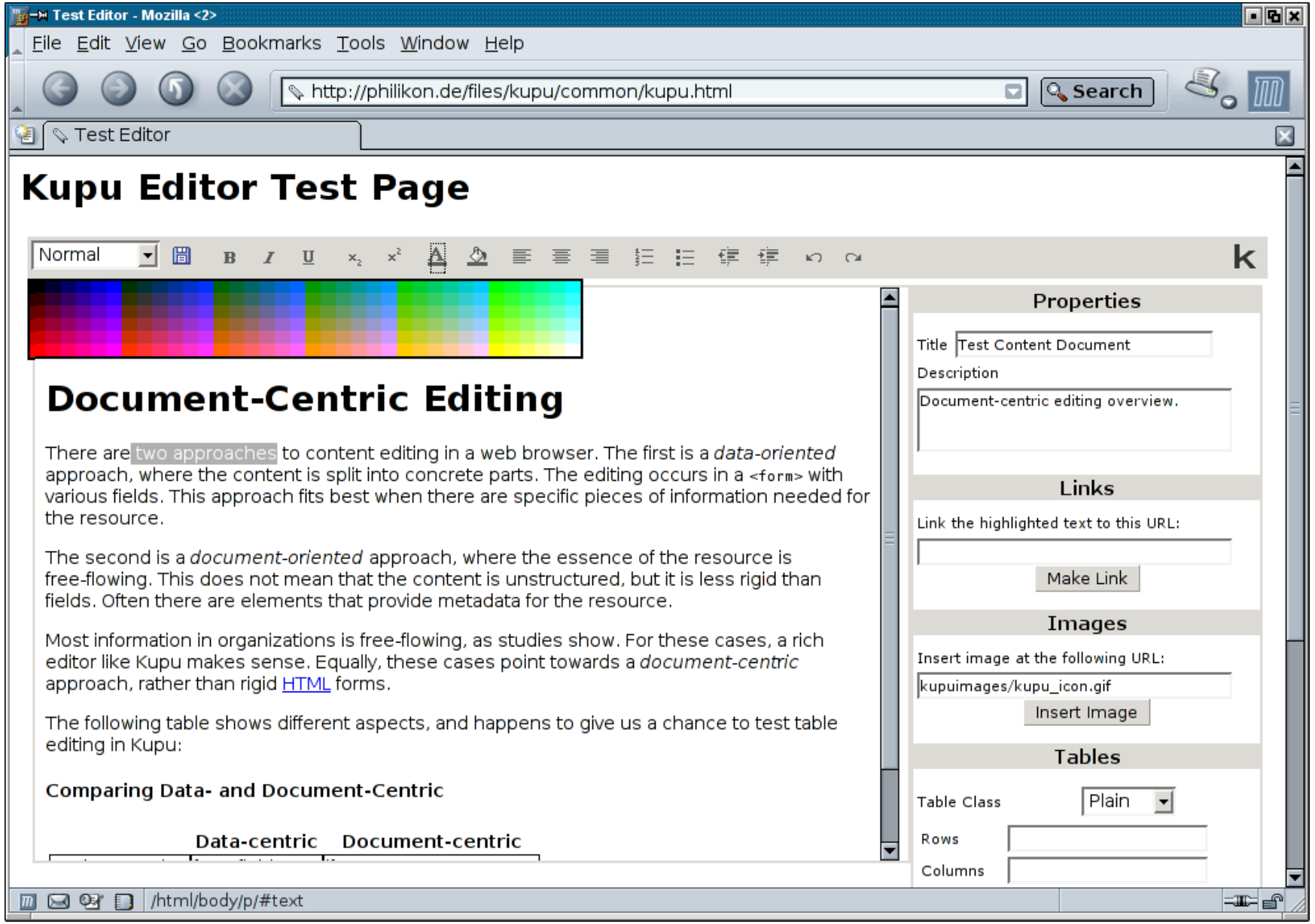


"from clay tablets to PDFs, we have slowly evolved various models of 'a document' that have many features in common … this could not have been done without there being a mental model of the document to start from"

– Peter Flynn

A Grammar of Content Models

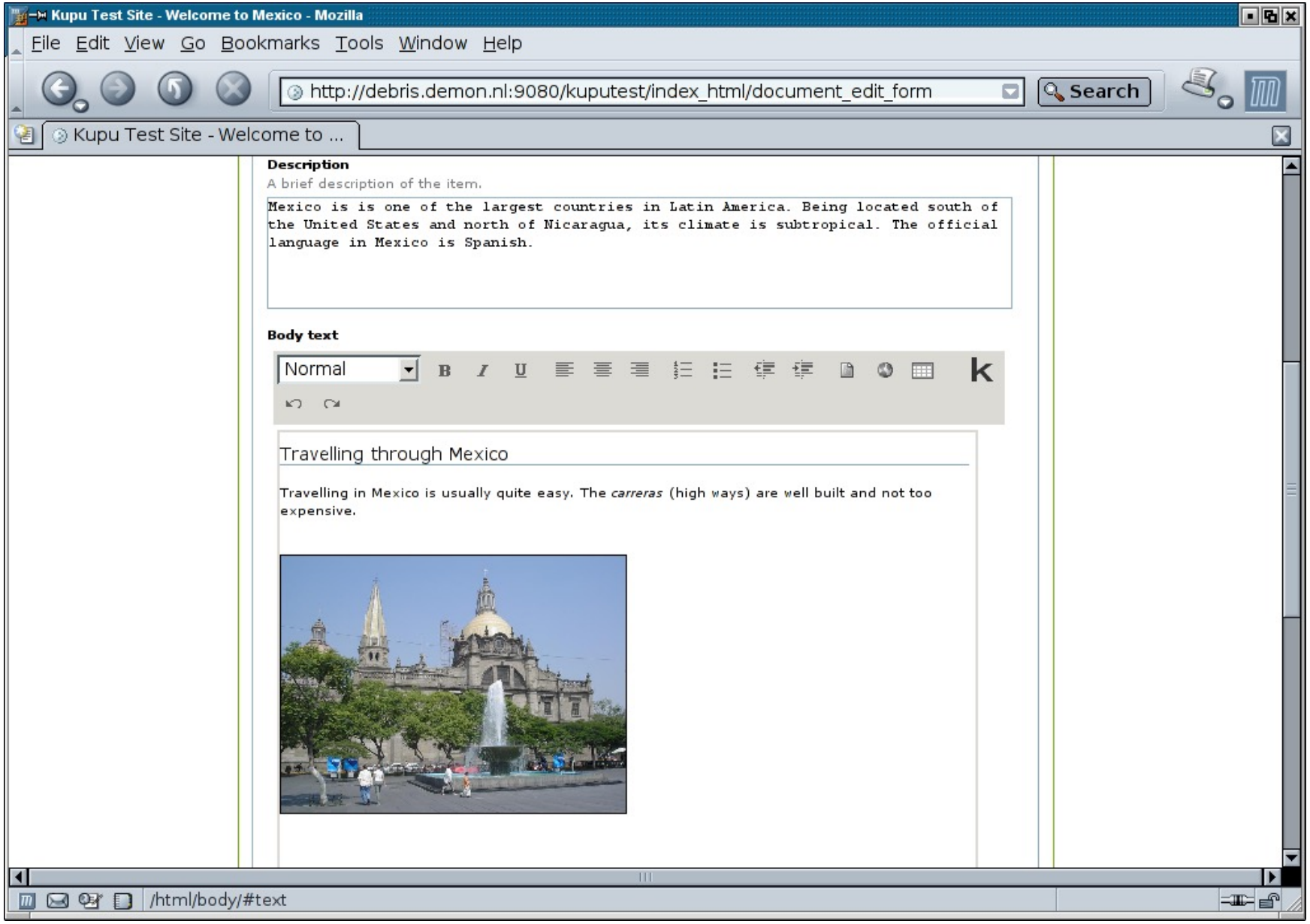
Early WYSIWYG Editors



Kupu editor test page, circa 2004

Marc Salvatierra

A Grammar of Content Models



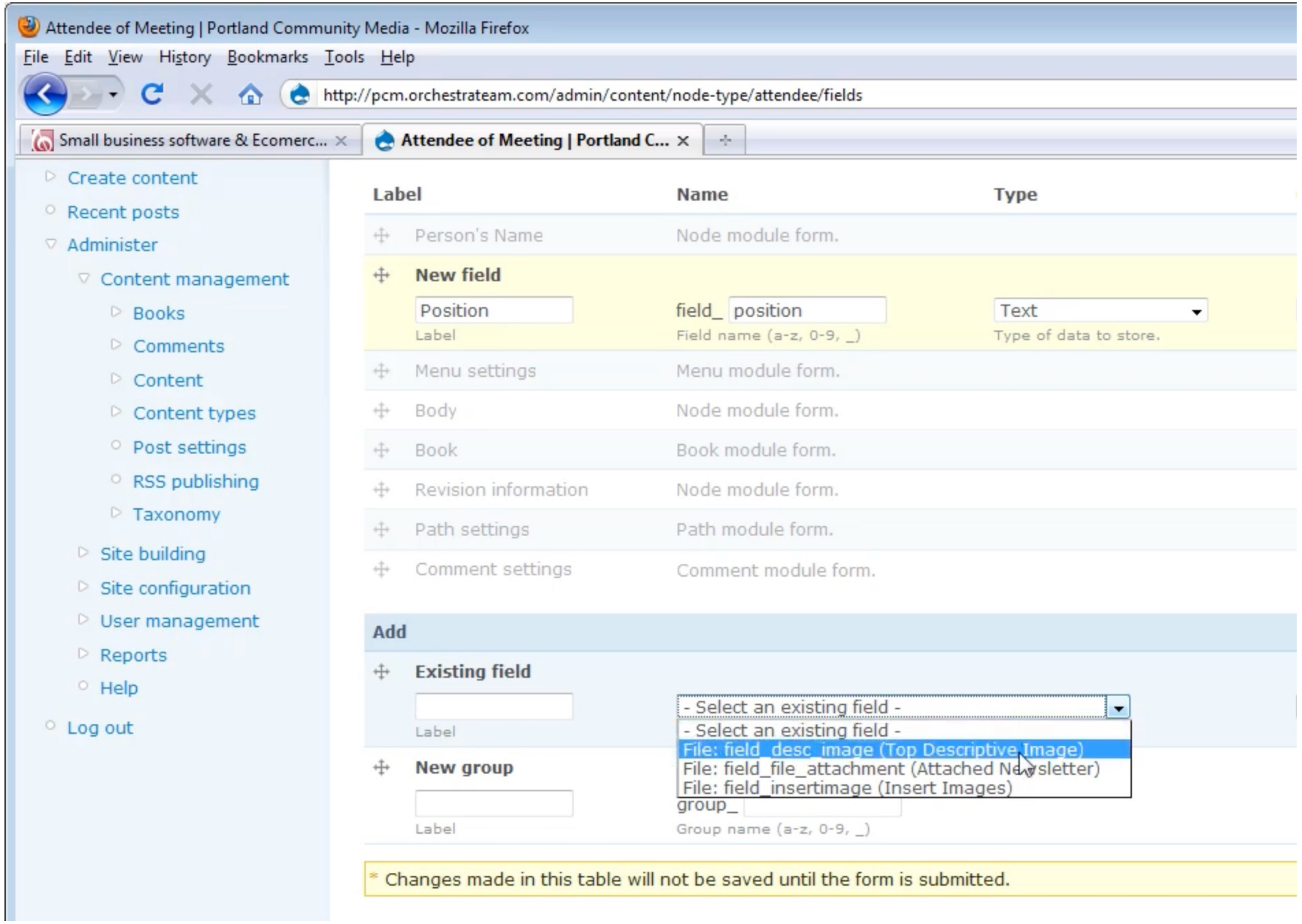
Early WYSIWYG Editors

Kupu editor in Plone document, circa 2004

Marc Salvatierra

A Grammar of Content Models

Early Configuration

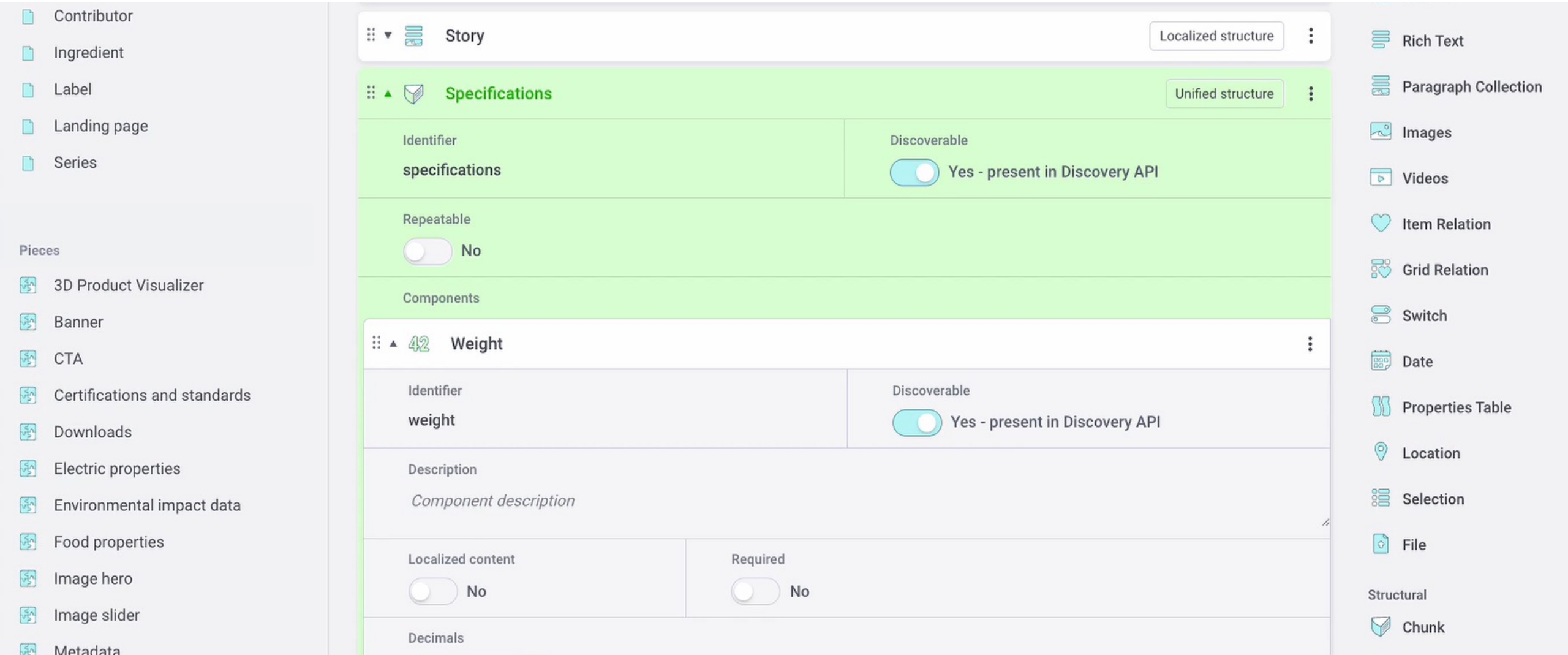


Drupal
Content Construction Kit – CCK
OrchestraTeam, circa 2010

Marc Salvatierra

A Grammar of Content Models

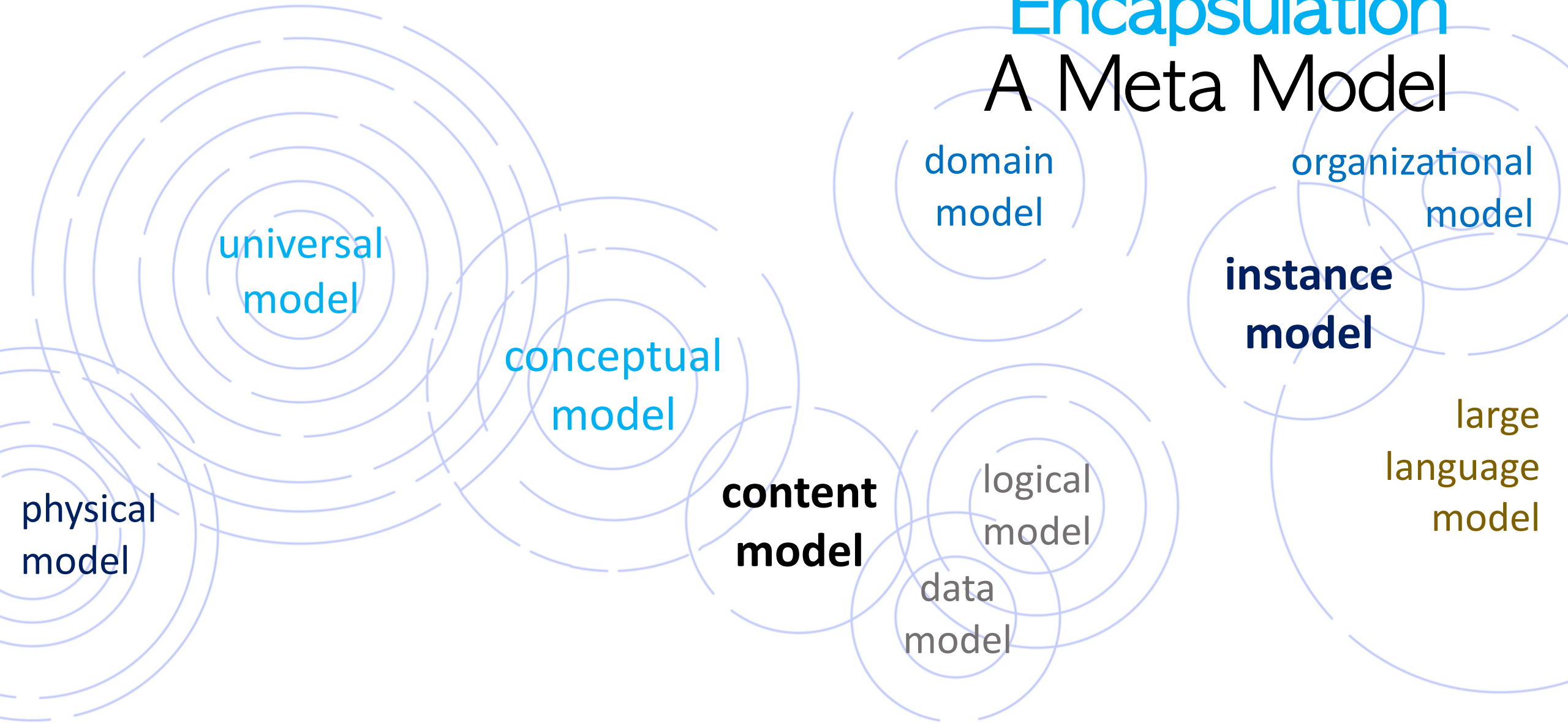
‘Shapes’ Modeling in Crystallize PIM



Marc Salvatierra

Encapsulation

A Meta Model



Encapsulation

The Content "Middle Model"

Content Model Centrality – The current CMS content model sits among larger, smaller, outer and inner models.

External Models – We already have models like ontologies, org models and domain models outside the content model.

Internal Models – Schema is page level. The smaller interior fragments of field-level content instances are less modeled.

Expanding the Model

Authoring in the Large & Small

Origins – *Programming-in-the-Large versus Programming-in-the-Small* (DeRemer & Kron, 1976) … *HDM – A Model for the Design of Hypertext Applications* (Garzotto, Paolini & Schwabe, 1991)

Distinction – Managing content across a schema-based content model is authoring in the large; populating instances in text, rich text, WYSIWYG and other fields and slots is authoring in the small.

Parallel Approach – *Modeling in the Large and Modeling in the Small* (Jean Bézivin and the ATLAS group, 2005)

Expanding the Model

Content Modeling in the Large

Motivation – Your content model reaches for meaning and integration outside the standard content management system.

Linked Data – Semantics, graphs and ontologies now operate at this broad level, beyond Schema.org markup, into inference.

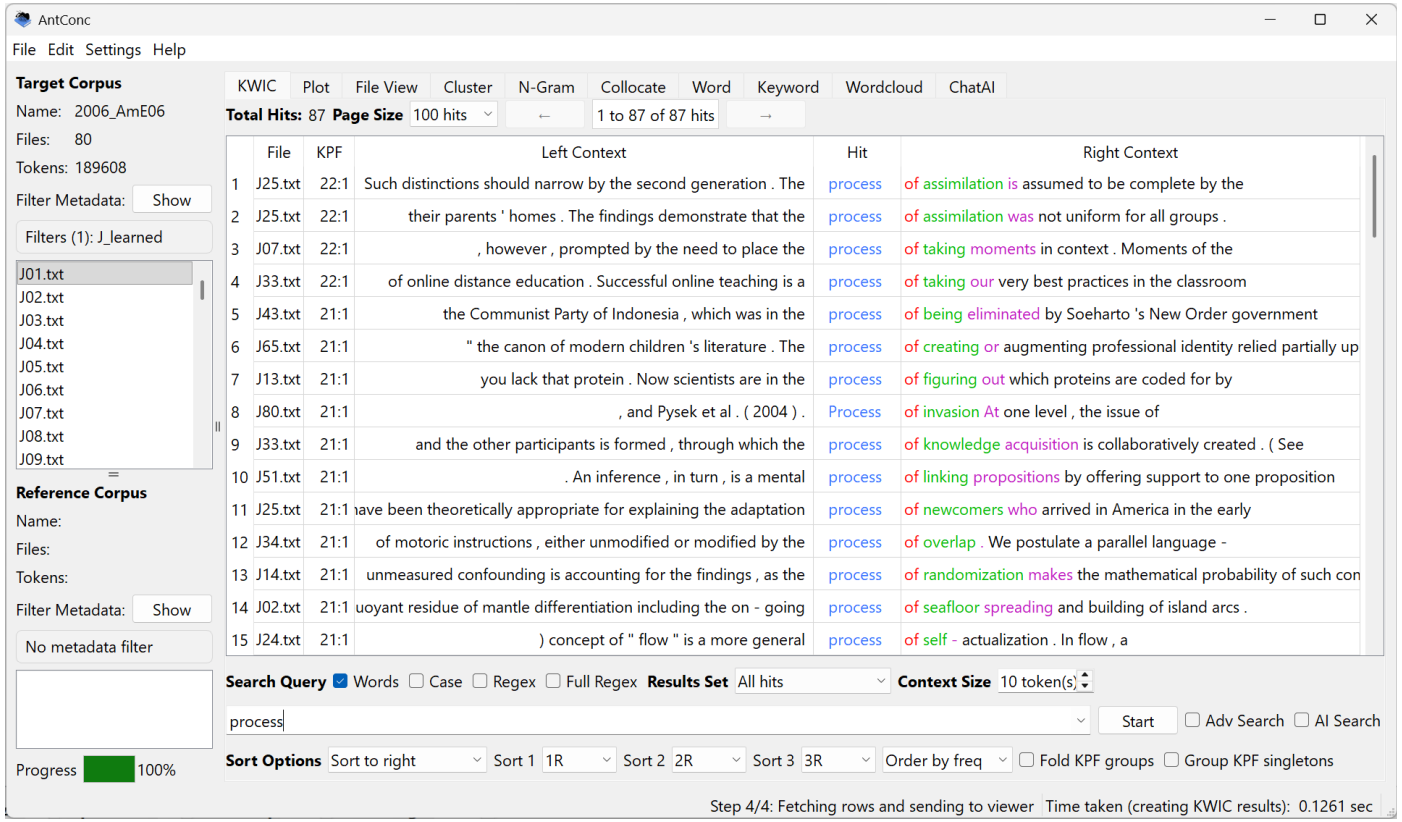
Megamodeling – Building a navigable ecosystem of models, metamodels, their derivations and transformations. (Bézivin)

Marc Salvatierra

Content Modeling in the Large
Text As Data & Text Analytics

treats large-scale unstructured text as quantifiable data to extract patterns, topics, entities; employs NLP, ML, LLMs, etc. (computational linguistics, 1950s – present)

AntConc – a freeware corpus analysis toolkit for concordancing and text analysis (Laurence Anthony)

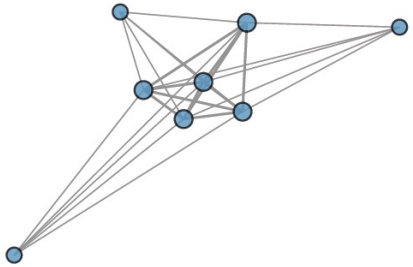


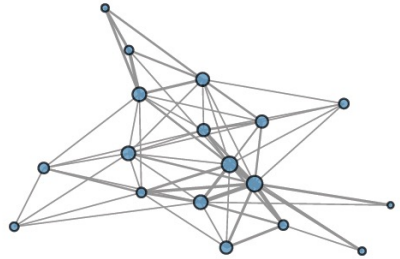
https://www.laurenceanthony.net/software/antconc/

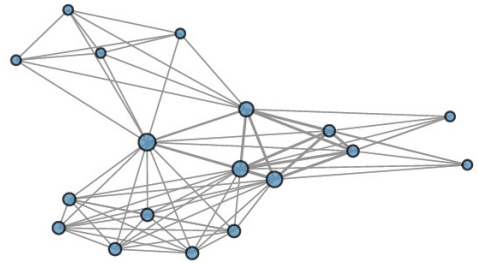
Content Modeling in the Large
Distant Reading

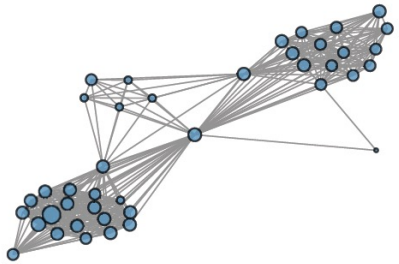
method for discovering large patterns across corpora in literary studies and digital humanities; employs text mining, NLP, network analysis, etc. (Franco Moretti, 2000)

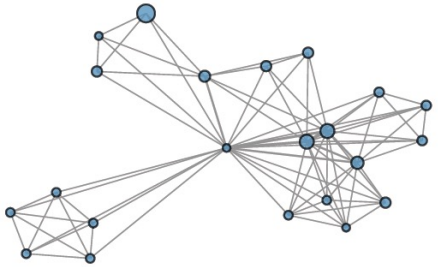
character networks in German dramas, by author

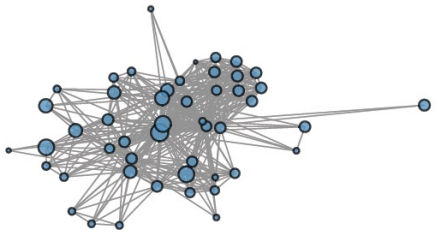

Dehmel, DM, 1917


Hofmannsthal, DS, 1917


Wildgans, Di, 1918


Kaltneker, DO, 1925

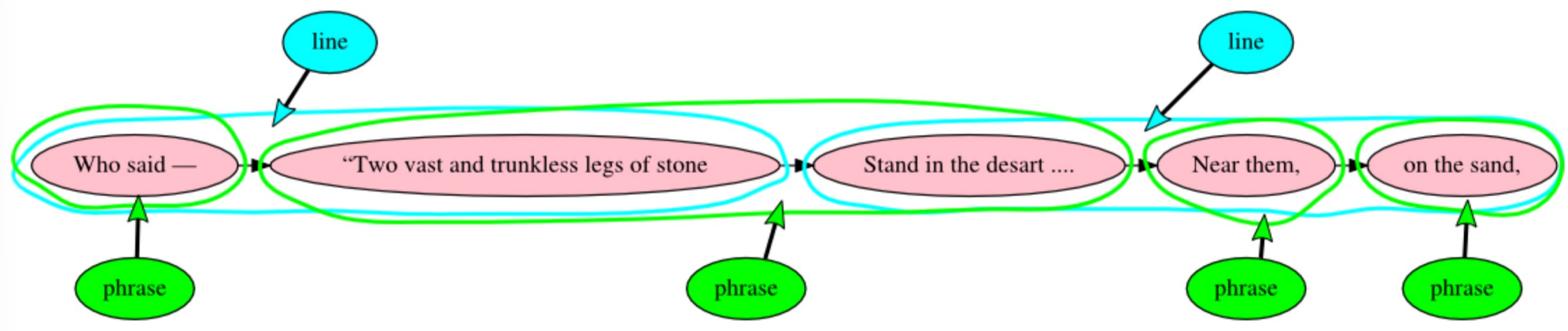

Klabund, DK, 1925


Mühsam, S, 1928

Frank Fischer, "'Distant-Reading Showcase': Designing Our DHd2016 Conference Poster," DLINA (blog), March 30, 2016, https://dlina.github.io/Distant-Reading-Showcase-Poster-DHd2016-Leipzig/

Content Modeling in the Large
Text As Graph & TAGML

hypergraph-based text-modeling approach to overlapping, discontinuous, multiple hierarchies; used in digital humanities, scholarly editing (Huygens Institute, 2017 – 2018)



hypergraph showing overlap of phrases and metrical lines in Percy Bysshe Shelley’s “Ozymandias”

Ronald Haentjens Dekker and David J. Birnbaum, "It’s more than just overlap: Text As Graph," in *Proceedings of Balisage: The Markup Conference 2017*, Balisage Series on Markup Technologies, vol. 19 (2017), https://doi.org/10.4242/BalisageVol19.Dekker01

Content Modeling in the Large

The House Model

"not a generic [AI] model pointed at internal documents, but one that 'learns'" from org data, processes, decisions (Matt Robinson, *The Rise of the House Model*, 2026)

"There should be as many models in the world as firms in the world."

– Microsoft CEO Satya Nadella

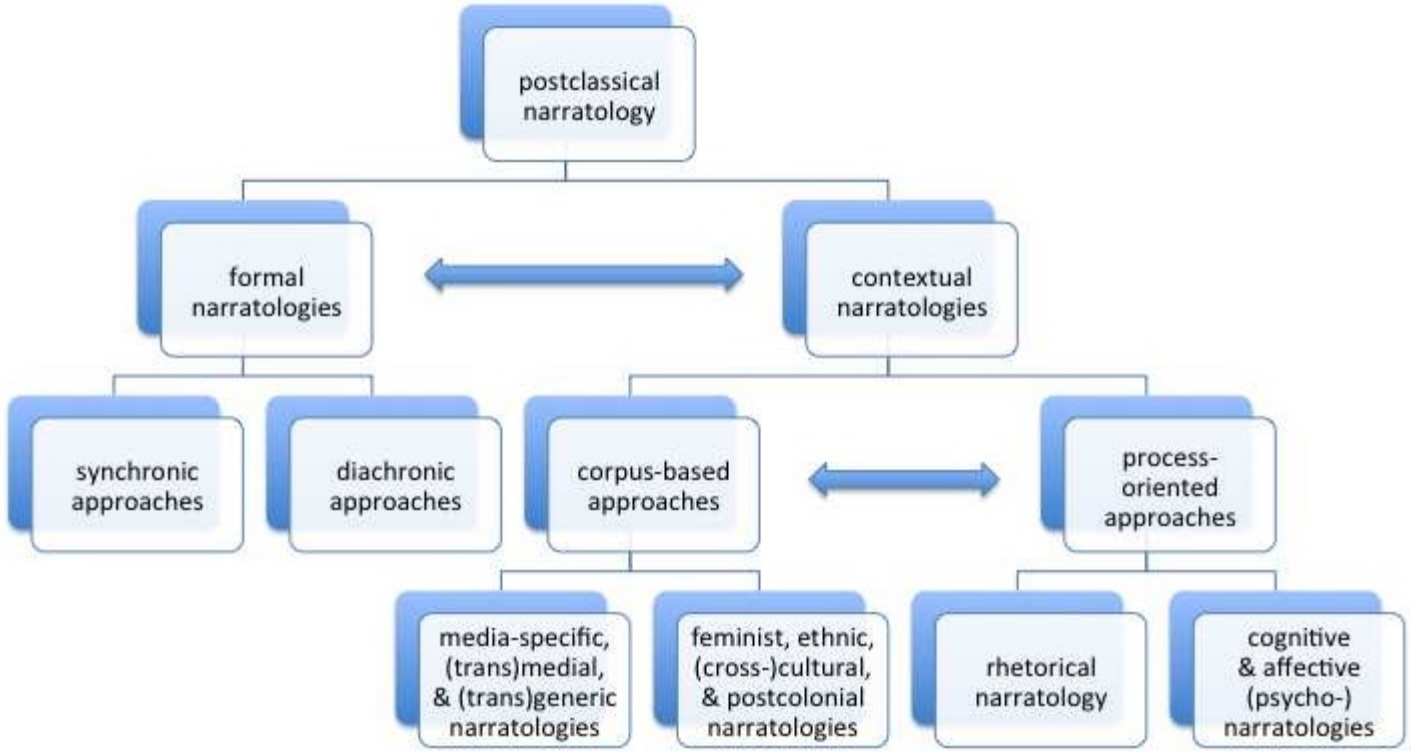


Matt Robinson, "The Rise of the House Model," *AI Street*, July 7, 2026, https://www.ai-street.co/p/the-rise-of-the-house-model

Content Modeling in the Large
Post-Classical Narratology

studies how narratives are structured, processed and experienced in context, across media; draws on cognitive science, rhetoric and cultural theory (late 1990s onward)

model presenting approaches in post-classical narratology

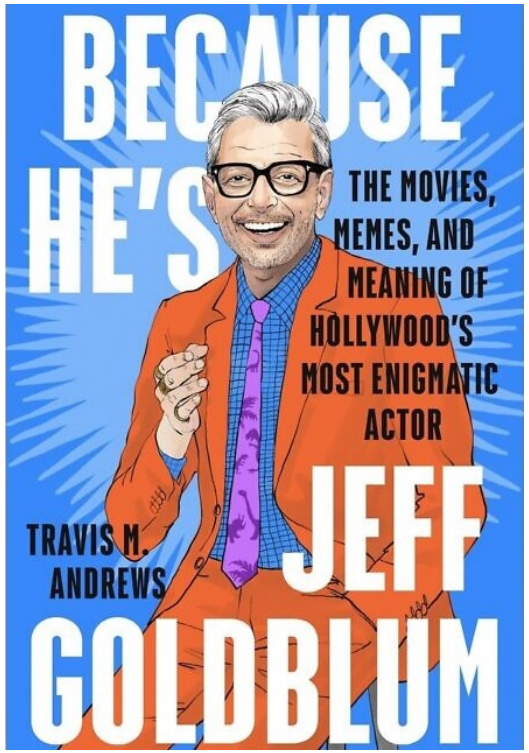


Roy Sommer, "The Merger of Classical and Postclassical Narratologies and the Consolidated Future of Narrative Theory," *DIEGESIS* 1, no. 1 (2012): 153

Content Modeling in the Large

Peritext

text elements in an instance that guide its reception: dedication, preface, epigraph, etc.; used in literary theory, bibliography, translation studies (Gérard Genette, 1987)



10. Jeff Goldblum vs. the Idea of Jeff Goldblum
235

INTERLUDE #10: The Recipe for Jeff Goldblum's Instagram Account
259

11. A Living Meme
261

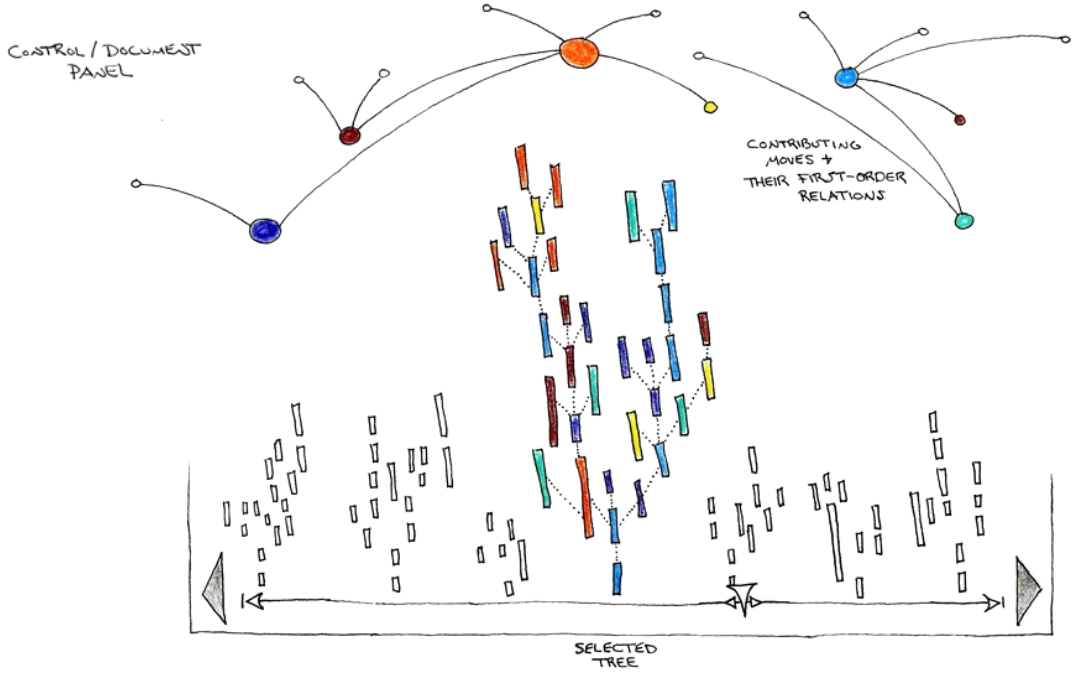
EPILOGUE #1 (& ONLY): Life Finds a Way and Goldblum Finds a Life (i.e., the Chapter We "Need")
287

Travis M. Andrews, *Because He's Jeff Goldblum: The Movies, Memes, and Meaning of Hollywood's Most Enigmatic Actor* (New York: Plume, 2021)

Content Modeling in the Large
Radiant Textuality

theory of texts as variable, multi-dimensional networks; applied in digital humanities, textual criticism, scholarly editing; uses SGML / XML, TEI (Jerome McGann, 2001)





sketches for *IVANHOE*, a collaborative platform for textual interpretation

Left: a control/document panel for players' roles, moves and rationales. Right: a 'selected tree' of moves showing contributing moves and their first-order relations. Bethany Nowviskie, "sketching ivanhoe," Bethany Nowviskie (blog), June 5, 2009, https://nowviskie.org/2009/sketching-ivanhoe/

Expanding the Model

Content Modeling in the Small

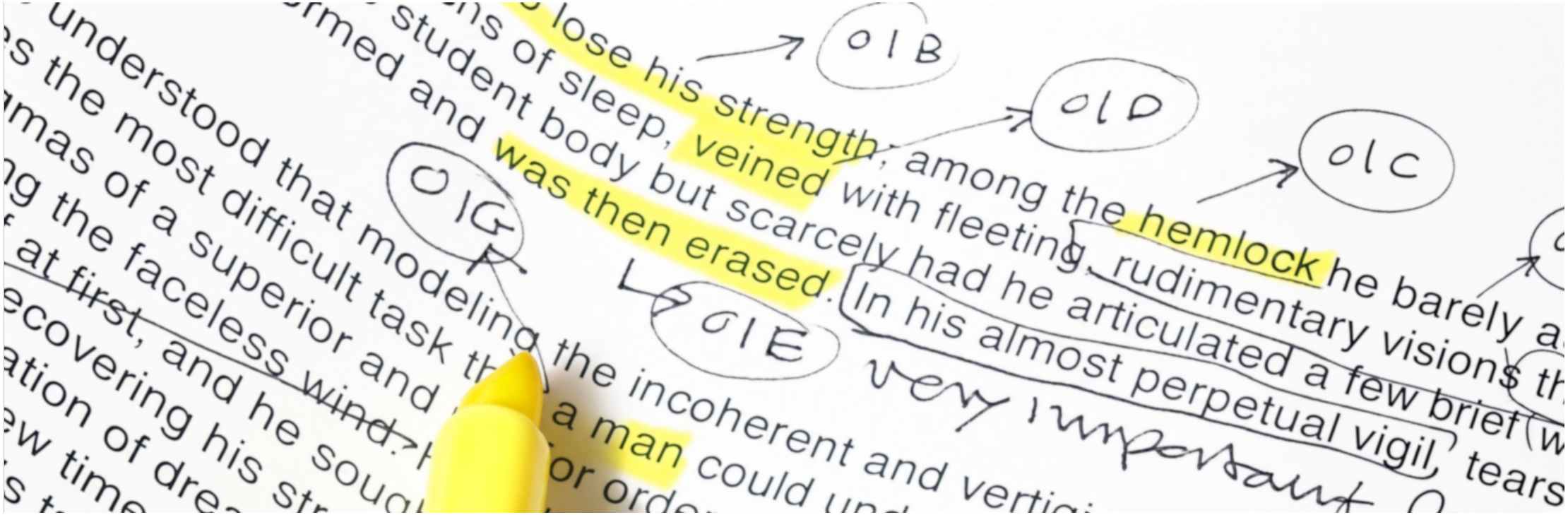
Instance-Level Crisis – Jeff Eaton remarked on the *Battle for the Body Field* and being *Buried in Blocks;* Karen McGrane spoke of the "war of Blobs versus Chunks".

Modeling Mystery – LLMs use content from our pages, embedded in segments and snippets – but stripped of structure. What meaning these LLMs derive and by what mechanics is opaque.

Micromodeling – Digital humanities approaches like the Text Encoding Initiative (TEI) micromodel content at the inline phrase and string level, beyond semantic HTML and document grammar.

Content Modeling in the Small
Close Reading

asserts texts hold implicit value; detailed dissection of language, structure, form; used in literature, humanities, legal analysis (New Criticism, practical criticism)



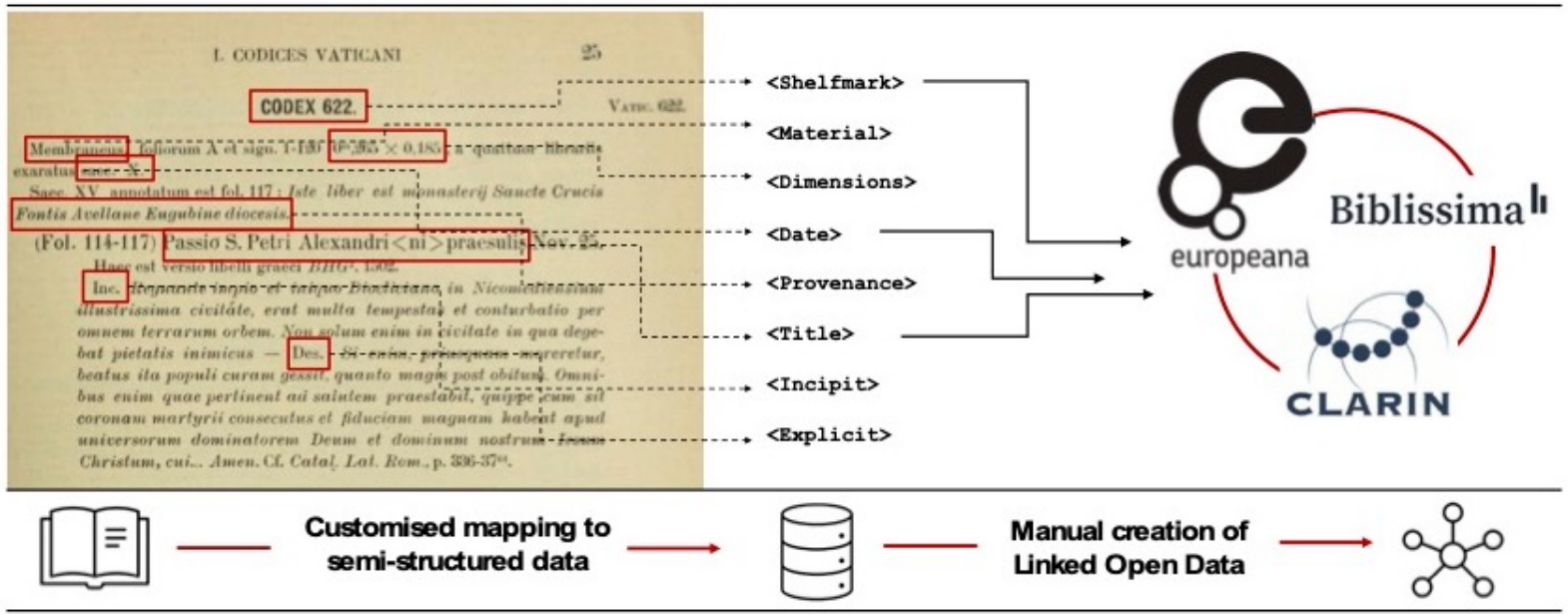
Guido "random" Alvarez, CC BY 2.0 <https://creativecommons.org/licenses/by/2.0>, via Wikimedia Commons – edited in Grok

Content Modeling in the Small

Semantic Annotation

attaching machine-readable meaning to content; used in knowledge-intensive systems; employs RDF, OWL, XML / TEI, RDFa, JSON-LD, Schema.org, W3C WADM

semantic annotation of a Vatican catalogue entry in Manuscript-AI ontology



Shari Boodts and Giulia Biagioni, "Manuscript-AI: A Dedicated Ontology and AI-Tool to Facilitate Integration and Accessibility of Heterogeneous Catalogue Data on Medieval Manuscripts" (presentation, 60th Joint Meeting of the CIDOC CRM SIG and 53rd LRMoo SIG, Universität Bern, Bern, Switzerland, April 3, 2025), https://cidoc-crm.org/sites/default/files/CIDOC_Presentation%20Manuscript%20AI.pdf

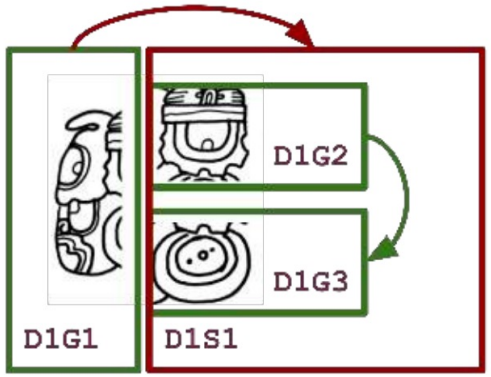
Content Modeling in the Small

Text Encoding Initiative (TEI)

standard for encoding machine-readable texts; used in digital humanities, libraries, scholarly editing; employs XML, schemas, ODD, XSLT (ACH, ALLC, ACL, 1987)

TEI encoding of relative positions of signs in a Maya glyph block

```
<tei:ab xml:id="D1">
    <tei:g xml:id="D1G1" tei:rend="left_beside"
    tei:corresp="#D1S1"/>
    <tei:seg xml:id="D1S1" tei:rend="right_beside"
    tei:corresp="#D1G1"/>
            <tei:g xml:id="D1G2" tei:rend="above"
            tei:corresp="#D1G3"/>
            <tei:g xml:id="D1G3" tei:rend="beneath"
            tei:corresp="#D1G2"/>
    </tei:seg>
</tei:ab>
```



Martin de la Iglesia et al., "The Code of Maya Kings and Queens: Encoding and Markup of Maya Hieroglyphic Writing," *Journal of the Text Encoding Initiative*, no. 14 (2021), https://doi.org/10.4000/jtei.3336

Content Modeling in the Small
Literary Annotation

attaching notes, glosses, commentary to texts; used in literary scholarship, education, digital humanities; uses TEI-XML markup, standoff annotation, W3C Web Annotation

tropological encoding of John Keats' "Ode on a Grecian Urn"

```
<theme type="immortality">
    <l n="1"><litFigure type="apostrophe">THOU</litFigure> <ambiguity
    item1="immobile" item2="continuation">still</ambiguity>
    <negation>un</negation><connotation item="aggression"
    value="neg">ravish'd</connotation><litFigure type="metaphor"
    tenor="urn">bride</litFigure> of <w reg="silence">quietness</w>, </l>

    <l n="2"><litFigure type="apostrophe">Thou</litFigure>
    <connotation item="adopted" value="neg">foster-</connotation>
    <litFigure type="metaphor" tenor="urn">child</litFigure> of <litFigure
    type="personification"><w reg="silence">Silence</w>
    </litFigure> and slow <litFigure type="personification">Time</litFigure>, </l>
</theme>
```

Kate Singer, "Digital Close Reading: TEI for Teaching Poetic Vocabularies," *The Journal of Interactive Technology and Pedagogy*, no. 3 (May 2013), https://cuny.manifoldapp.org/read/digital-close-reading-tei-for-teaching-poetic-vocabularies-bc1c3823-8397-4973-a5fd-c2c1ea76172e/section/01dadc0e-715d-4027-b12e-6e5f9a278aac

Content Modeling in the Small

Linguistic Annotation

marking up linguistic structure: parts of speech, syntax, semantics, discourse; serves computational linguistics / NLP (corpus linguistics, 1960s; systematic annotation, 1970s)

Brown Corpus tagging: ‘“Open Source CMS 26” conference is happening in Utrecht.’

```
''/''  Open/JJ-TL  Source/NN-TL
CMS/NP-TL  26/CD-TL  ''/''
conference/NN  is/BEZ  happening/VBG
in/IN  Utrecht/NP  ./.
```

https://varieng.helsinki.fi/CoRD/corpora/BROWN/tags.html

Marc Salvatierra

Content Modeling in the Small

Instance-Level Modeling

We need CMS facilities to micromodel content at the internal field instance level, *à la* Plate Delta CMS' structuring assistant.

HTML need not be the only markup surface – we can swap and apply various markup regimes to any internal content instance.

Marked-up content is still subject to AI splitting, embedding and retrieval – a process we can influence but can't control.

Marc Salvatierra

Open World Content
New Directions in Content Modeling

Making Before Modeling – Humans create then formalize.

Presentation As Content – Aesthetics are always included.

Abundance, Bustle, Collage – Organic balances organized.

Self-Description – Content should convey its context.

Marc Salvatierra

New Directions in Content Modeling

Making Before Modeling

Humans tend to publish content before methodically examining its essence; later re-encounters inform structure.

Creation often precedes or bypasses formalization, as seen in ephemeral information objects and organic aggregations.

We should indulge this reality, leaving the canvas blank and humans free to experiment outside pre-set limits.

New Directions in Content Modeling

Presentation As Content

Separation of presentation from content privileges text over contentual aesthetics that also convey meaning and effect.

The content of zines and Digital-Out-of-Home installations is highly presentational in generation and consumption.

Art and craft still have their rightful place in content modeling, leading consumers deeper into experience and engagement.

New Directions in Content Modeling

Abundance, Bustle & Collage

Humans eventually shift to wanting to consolidate content into minimal models and sterile systems – to increase their control.

Yet, "clutter" can still carry a culturally appropriate sense of activity, excitement and life: Kowloon, Shinjuku, Times Square.

Ad-hoc aggregation transmits valuable data – content on a neighborhood café bulletin board remains part of a local model.

Marc Salvatierra

New Directions in Content Modeling

Self-Described Content Instances

A book with no title page in a library catalog is fine until it leaves the library; conversely, libraries need title pages to build catalogs.

We can publish the content model within, among or alongside instances, projecting a *decompilable, repurposable content model*.

Content instances can collectively convey elements of the model to which they belong – projecting a *distributed content model.*

Marc Salvatierra

Conclusion
Reflections

“The Model” Is Meta – Our CMS content model can inhabit, contain and connect with many other models.

Expand the Content Model – Bridge and blend the content model with other models; model interior content instances in granularity.

Model Content in New Directions – After creation … up to ontologies … down to strings … in to self description, out to presentation … and together in aggregation.

Marc Salvatierra

Thank You

