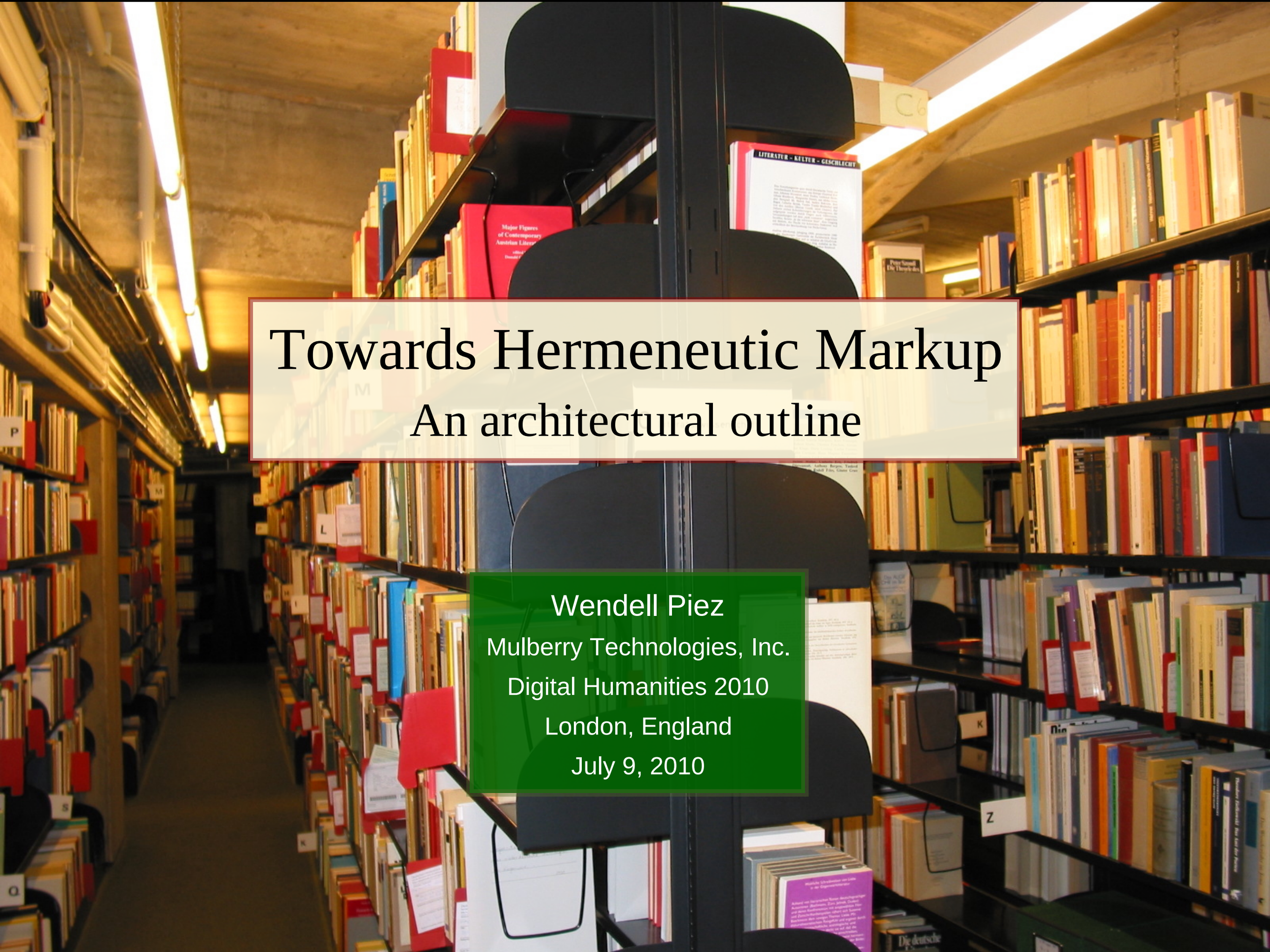


Towards Hermeneutic Markup

An architectural outline

Wendell Piez
Mulberry Technologies, Inc.

Digital Humanities 2010
London, England
July 9, 2010

What ~~Is~~ *would be* “Hermeneutic Markup”?

Interpretive
Expressive
Responsive
Idiosyncratic
Peculiar
Open-ended

Oriented towards *process* not *product*

(Almost) An Example

Mark up a sonnet – twice
Once with verse structure
Once for grammar/phrasing
– then merge them –

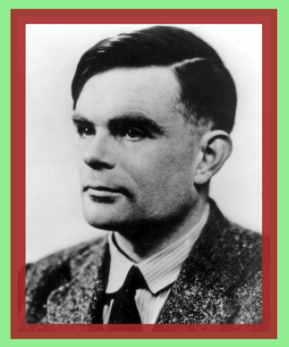
```
<sonnet>
  <octave
    <quat
      <li
      <li
      <li
      <li
    </qua
    <quat
      <li
      <line>The feathered glory from her
      <line>And how can body, laid in th
      <line>But feel the strange heart b
    </quatrain>
  </octave>
  <sest
    <te
      <
      <
      <
    </t
    <te
      <
      <
      <
    </t
  </ses
</sonn
```

```
<sonnet><s><phr>A sudden
Above the staggering g
By the dark webs,</p
<phr>He holds her
<s><phr>How can those terrified
The feathered glory from her loo
<s><phr>And how can body,</phr>
<phr>But feel the strange heart
<s><phr>A shudder in the loins e
The broken wall,</phr> <phr>the
And Agamemnon dead.</phr></s> <s
<phr>So mastered by the brute bl
<phr>Did she put on his knowledg
Before the indifferent beak could let her drop?</ph
```

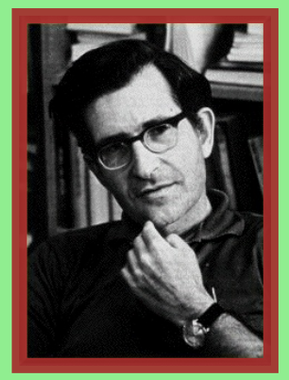
```
beating still</line>
thighs caressed</line>
```

So why only “almost”?
Merging them isn't so easy:
XML doesn't do overlap

How much do you know?



Turing



Chomsky

von Neumann

?

Melanie Carlstad

Abby Carlstad

Should you have to be an expert in formal languages to use markup technologies successfully?

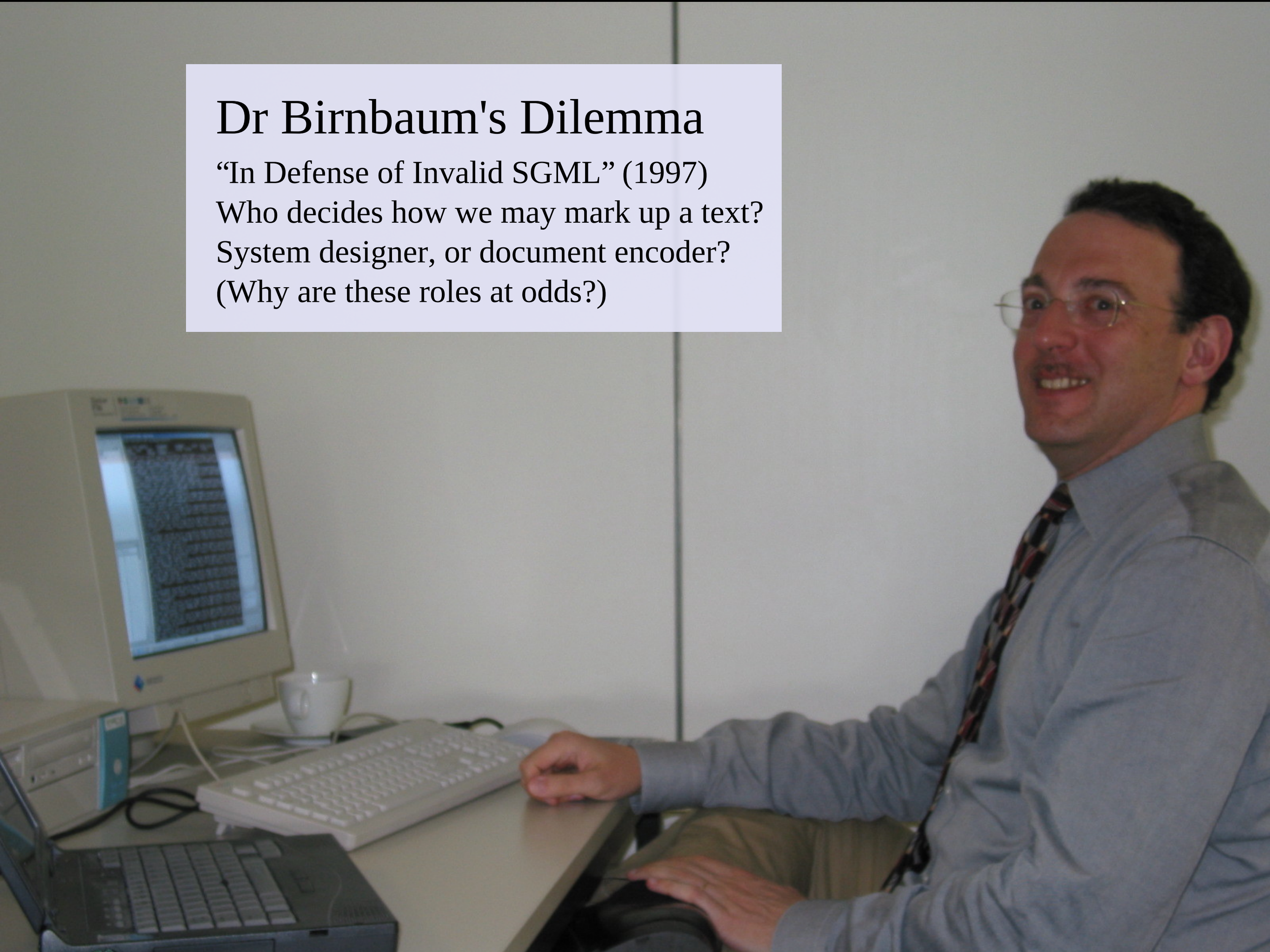
Like an iceberg

How big is it?
How close are we?
How much do we see?

Photo by Dr. Robert Ricker, NOAA
http://www.photolib.noaa.gov/htmls/rikr0119.htm

Dr Birnbaum's Dilemma

“In Defense of Invalid SGML” (1997)
Who decides how we may mark up a text?
System designer, or document encoder?
(Why are these roles at odds?)



Anomalies are interesting

In the real world, texts are mostly, but not *entirely*, regular in structure —

Snapshot of *Sartor Resartus*, by Thos. Carlyle
(Boston: James Munroe and co., 2nd ed. 1834)
Copy in the University of California Library
Scanned by Google Books (books.google.com)

Call this the “fractal geometry of information”

(after Benoit Mandelbrot, *The Fractal Geometry of Nature*)

An ironic man, with his sly stillness, and ambuscading ways, more especially an ironic young man, from whom it is least expected, may be viewed as a pest to society. Have we not seen persons of weight and name, coming forward, with gentlest indifference, to tread such a one out of sight, as an insignificancy and worm, start ceiling-high (*balkenhoch*), and thence fall shattered and supine, to be borne home on shutters, not without indignation, when he proved electric and a torpedo!”

Alas, how can a man with this devilishness of temper make way for himself in life; where the first problem, as Teufelsdröckh, too, admits, is “to unite yourself with some one and with somewhat (*sich anzuschliessen*)?” Division, not union, is written on most part of his procedure. Let us add, too, that, in no great length of time, the only important connexion he had ever succeeded in forming, his connexion with the Zähdarm family, seems to have been paralyzed, for all practical uses, by the death of the “not uncholeric” old Count. This fact stands recorded, quite incidentally, in a certain *Discourse on Epitaphs*, huddled into the present bag, among so much else; of which essay the learning and curious penetration are more to be approved of than the spirit. His grand principle is, that lapidary inscriptions, of what sort soever, should be historical rather than lyrical. “By request of that worthy nobleman’s survivors,” says he, “I undertook to compose his epitaph; and, not unmindful of my own rules, produced the following; which, however, for an alleged defect of Latinity, a defect never yet fully visible to myself, still remains unen-

graven;”—wherein, we may predict, there is more than the Latinity that will surprise an English reader.

HIC JACET

PHILIPPUS ZAEHDARM, COGNOMINE MAGNUS,

ZAEHDARMI COMES,

EX IMPERII CONCILIO,

VELLERIS AUREI, PERISCELIDIS, NECNON VULTURIS NIGRI

EQUES.

QUI DUM SUB LUNA AGEBAT,

QUINQUIES MILLE PERDICES

PLUMBO CONFECIT:

VARII CIBI

CENTUMPONDIA MILLIES CENTENA MILLIA,

PER SE, PERQUE SERVOS QUADRUPEDES BIPEDESVE,

HAUD SINE TUMULTU DEVOLVENS,

IN STERCUS

PALAM CONVERTIT.

NUNC A LABORE REQUIESCENTEM

OPERA SEQUUNTUR.

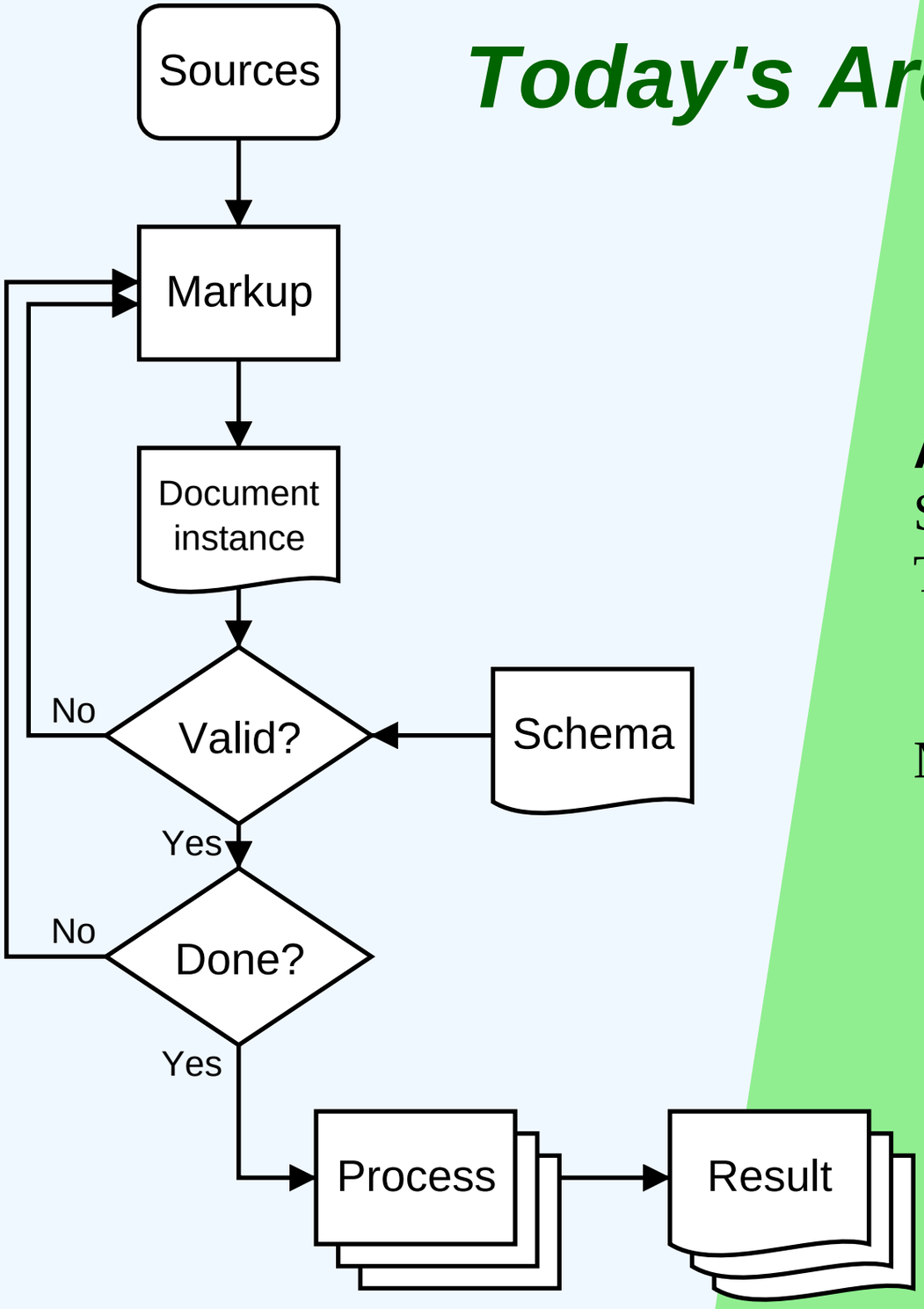
SI MONUMENTUM QUÆRIS

FIMETUM ADSPICE.

PRIMUM IN ORBE DEJECIT [*sub dato*];

POSTREMUM [*sub dato*].

Today's Architecture



Assumptions:

Schema (and processes) are complete and fixed

They are designed by experts with prior knowledge

– of significant textual features and phenomena

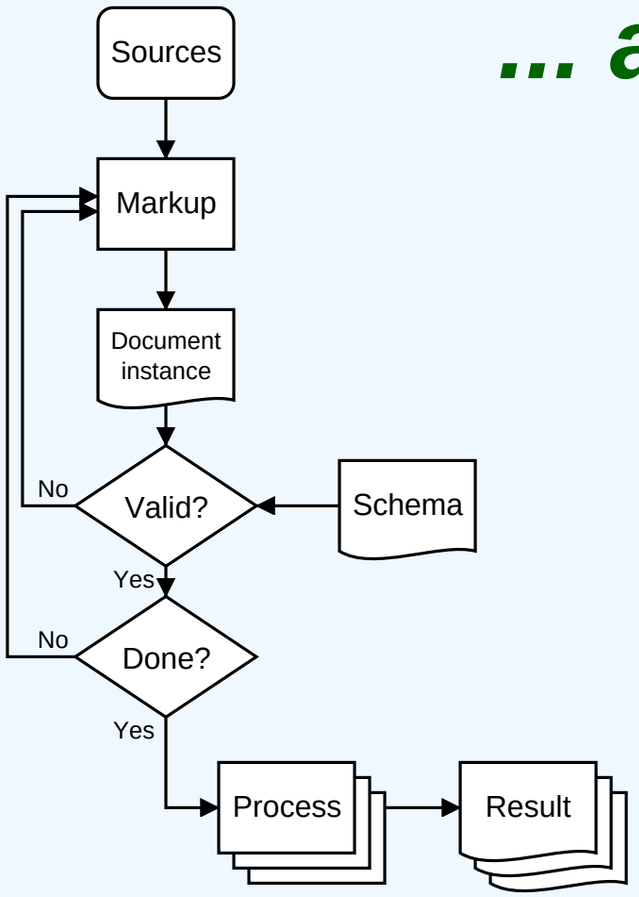
– of processing goals

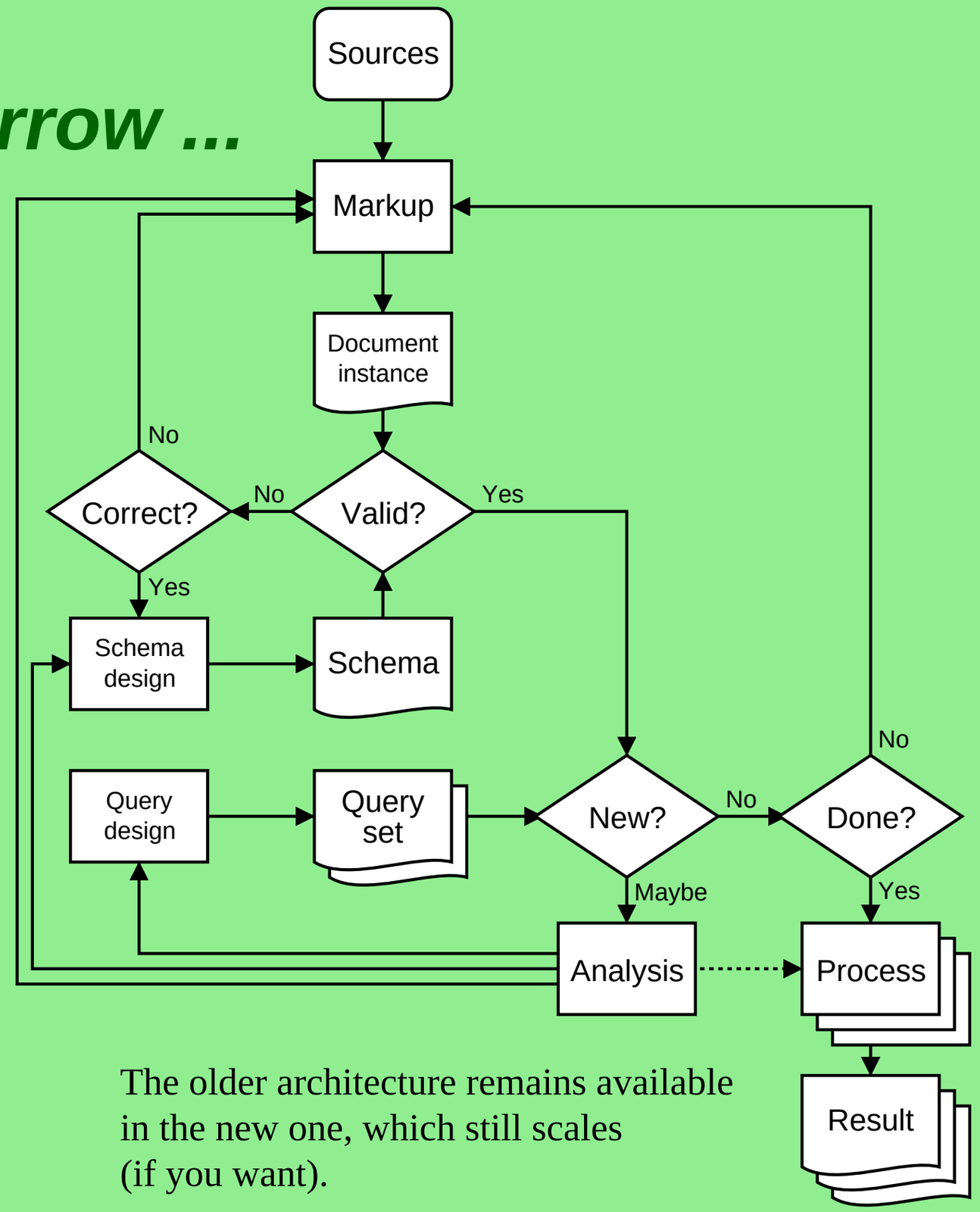
Markup serves the end of “production”

This is appropriate for publishing systems, whose primary requirement is to scale up to hundreds of thousands of texts of a given type, not to describe a single text *sui generis*.

Today ...

... and tomorrow ...





Thesis:
An interpretive “work” (edition, essay, monograph, schema, hypertext) is not a hermeneutic act, but the relic, trace, or residue of one.

The older architecture remains available in the new one, which still scales (if you want).

— *Desiderata* —

Openness and accessibility
Gentle learning curve
Build on current infrastructure

Data model supporting arbitrary overlap
One or more hierarchies, or none

Flexible and extensible standard schemas
Ad-hoc schemas
More than one schema per document

Extensible annotation framework
Annotations can be structured and tagged

Forms of scholarship enabled by markup

"Basis" (or "platform") editions, anthologies and collections
Scholarly editions (genetic, variorum, critical, etc.)
Networked annotation applications for research and classroom
Innovative modeling for "synoptic" and "hermeneutic" work
Allowing us to trace the evolution of textual structure

What is required:

External, non-proprietary ("standard") specifications
- Reasonably lightweight and learnable

A generalized data model that supports:
- Arbitrary (and "accidental") overlap
- Multiple concurrent hierarchies
- Layered, structured, tagged annotations
- Descriptive extensibility *ad hoc*

User interfaces for the model, including:
- A clean serialization syntax
- Various GUI widgetry
- Schema, transformation and query languages
- Stylesheets for known document types

Processes including:
- Query, transformation and conversion
- Schema validation and derivation
- XML induction/extraction

Promising developments:

Maturation of XML technologies
- Including querying, transformation, validation

Proliferation of freestyle annotation applications
- (You know who you are)

Convergence of initiatives for dealing with overlap
- MLCD (TexMECS/Goddag), LMNL, XStandoff, EARMARK and others

Emergence of digital humanities
- More widespread understanding of the issues