THE QUEST FOR CONTENT IN CONTEXT

Using standards and semantics for interoperability

Content, Seriously

OVERVIEW OF PRESENTATION

We will cover:

- The difference between data and content.
- Content and context.
- Content interoperability.
- Making more effective content, more effectively.
- Combining strengths of content and data.

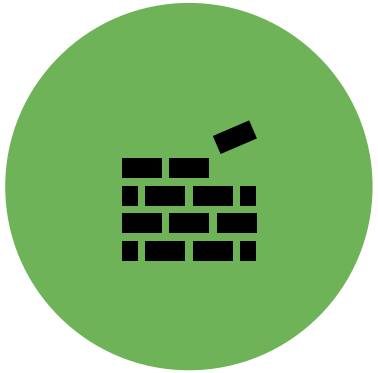
CONTENT VS DATA

Similarities and differences

DIFFERENCES BETWEEN CONTENT AND DATA

Narrative

Visualisation



Building blocks

UNDERSTANDING CONTENT

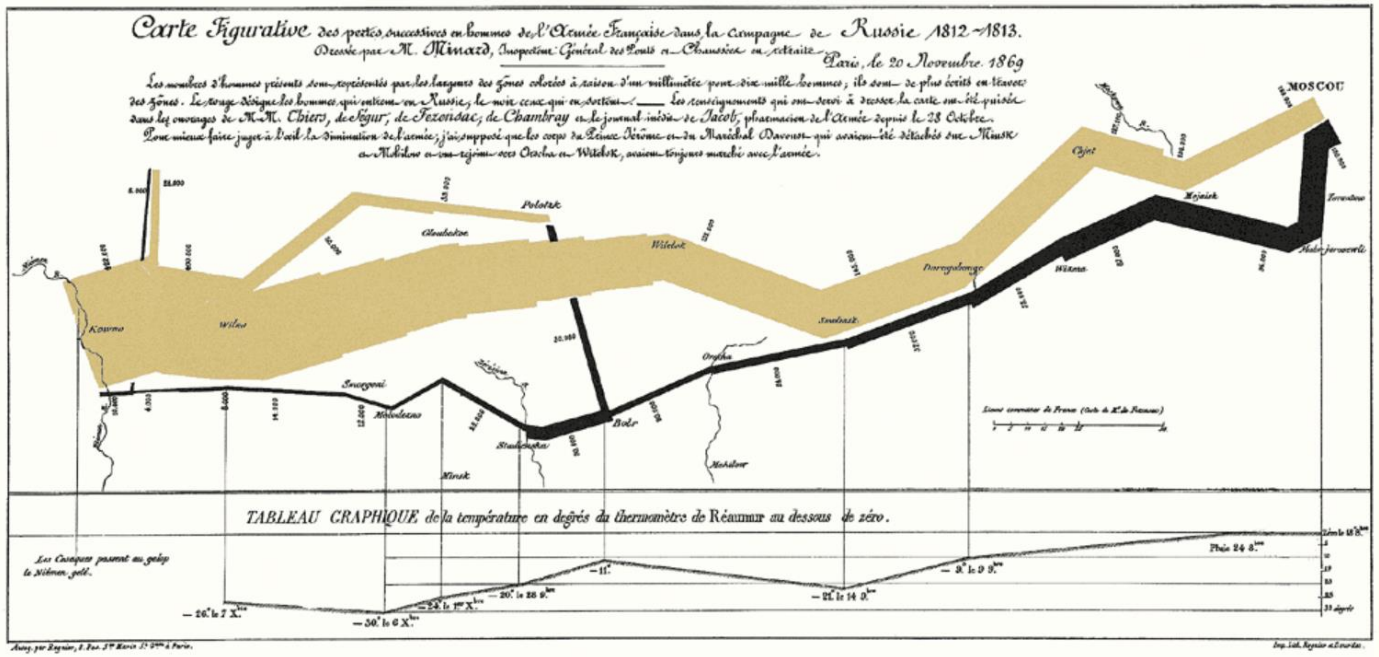
Models of text comprehension describe and explain the processes involved in understanding and remembering verbal information. (International Encyclopedia of the Social & Behavioral Sciences, 2001)

The band graph depicts Napoleon's army departing the Polish-Russian border. A thick band shows the size of his army at specific geographic points during their advance and retreat. The diagram displays 6 types of data in 2 dimensions: the number of Napoleon's troops, the distance travelled, the temperature, latitude and longitude, direction of travel, and location relative to specific dates.

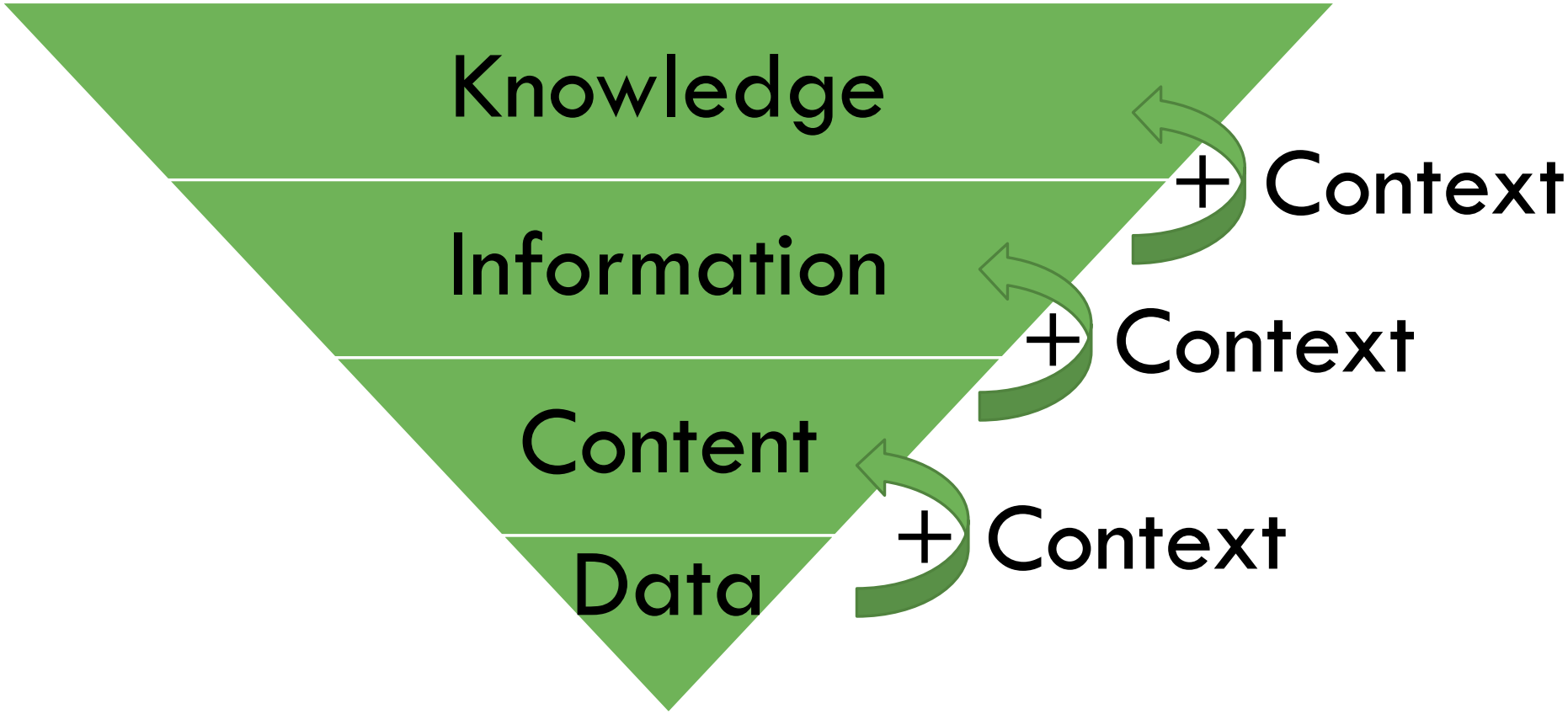
UNDERSTANDING DATA

Over 50% of study subjects had trouble interpreting statistical data, particularly when presented as probabilities instead of natural frequencies. (Science Daily, 2018)

Data is a visual medium – data visualisations. (Charles Joseph Minard, via Edward Tufte)



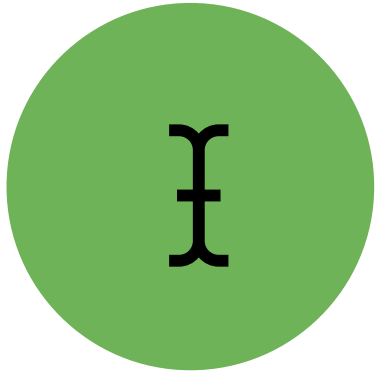
BUILDING BLOCKS TO KNOWLEDGE



CONTENT AND CONTEXT

If content is king, then context must be the empress

CONTENT AND CONTEXT



Context

Normalising and rationalising

Creating, enriching, context-building

CONTEXT IS EVERYTHING

What is the meaning of 42?

1. The ASCII symbol for *.
2. A data point.
3. A whole number between 41 and 43.
4. The answer to the ultimate question of life, the universe, and everything.
5. The total of my last grocery bill.
6. A week in October.

DATA MANIPULATION

Normalisation – structure a database to reduce redundancy and improve data integrity.

Standardisation – convert data to a common format to enable processing and analysis.

Rationalisation - creates/extends an ontology for a domain into the structured data world.

CONTENT MANIPULATION

Development – create content with relevance, accuracy, informativeness, timeliness, engagement, and editorial standards in mind.

Editing – structure a body of content to reduce redundancy and improve content integrity, including standardisation – rewrite content to fit common conventions to enable processing and analysis.

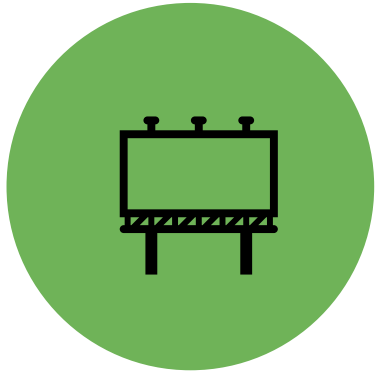
Reviews and approvals – submit to workflow, capture audit trail data, enforce editorial guidelines, and check quality control against standards.

Enrich - creates/extends an ontology for a domain into the structured content world.

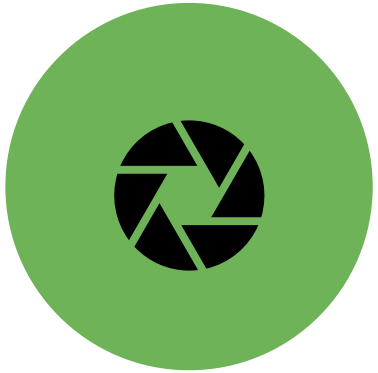
CONTENT INTEROPERABILITY

Schema and standards

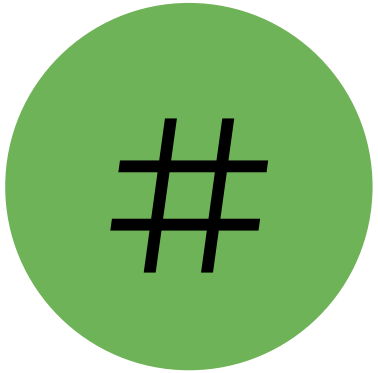
INTEROPERABILITY



Conventions



Standards



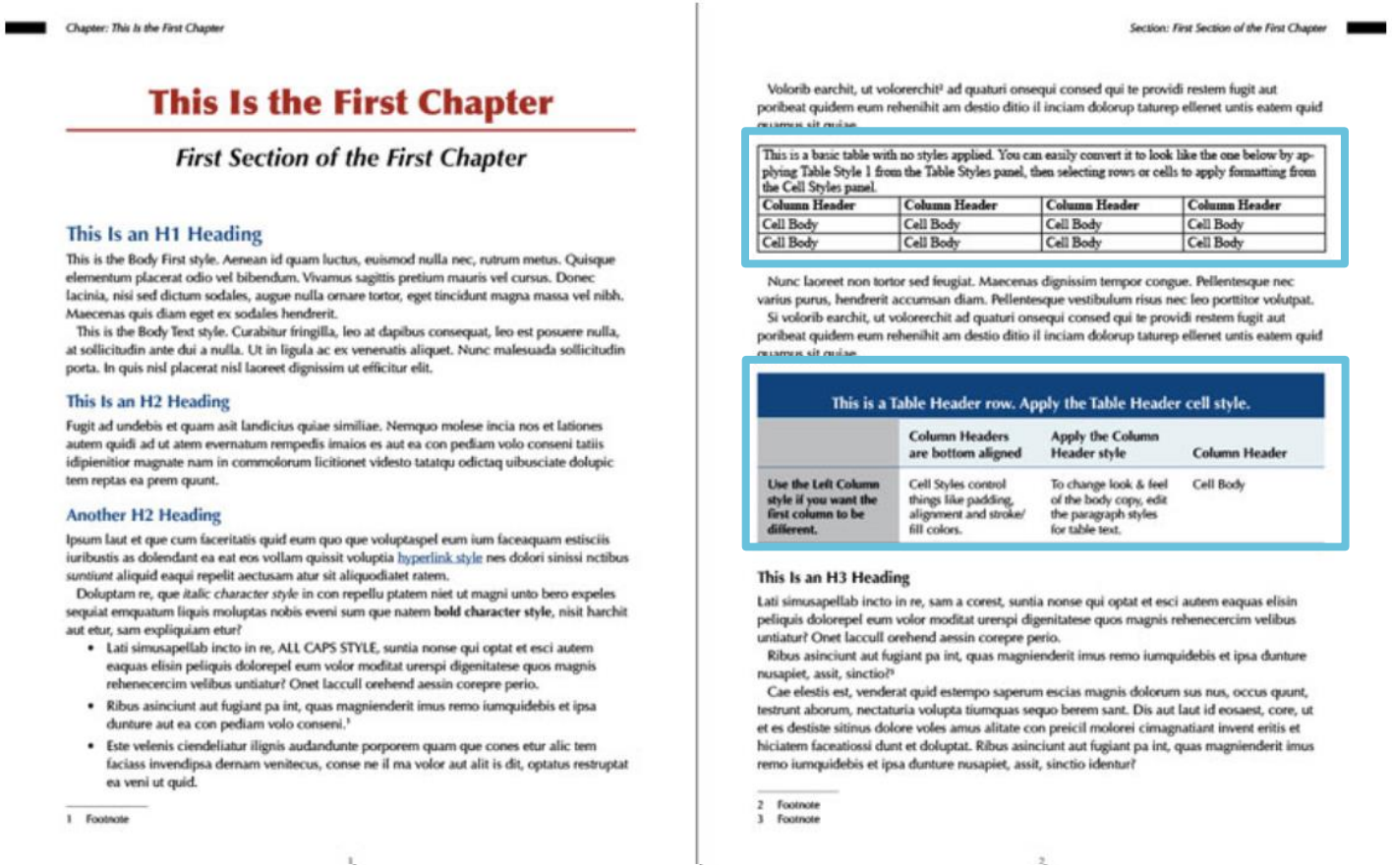
Semantics

CONVENTIONS ARE NOT STANDARDS

A convention is a usual way of presenting information.

Aids in comprehension.

Lots of discretion in expression, in placement, in attributes.



Chapter: This Is the First Chapter

This Is the First Chapter

First Section of the First Chapter

This Is an H1 Heading

This is the Body First style. Aenean id quam luctus, euismod nulla nec, rutrum metus. Quisque elementum placerat odio vel bibendum. Vivamus sagittis pretium mauris vel cursus. Donec lacinia, nisi sed dictum sodales, augue nulla ornare tortor, eget tincidunt magna massa vel nibh. Maecenas quis diam eget ex sodales hendrerit.

This is the Body Text style. Curabitur fringilla, leo at dapibus consequat, leo est posuere nulla, at sollicitudin ante dui a nulla. Ut in ligula ac ex venenatis aliquet. Nunc malesuada sollicitudin porta. In quis nisl placerat nisl laoreet dignissim ut efficitur elit.

This Is an H2 Heading

Fugit ad undebis et quam asit landicius quiae similiae. Nemquo molese incia nos et lationes autem quidi ad ut atem evernatum rempedis imaios es aut ea con pediam volo conseni tatiis idipienitior magnate nam in commolorum licitionet videsto tatatqu odictaq uibusciate dolupic tem reptas ea prem quunt.

Another H2 Heading

Ipsum laut et que cum faceritatis quid eum quo que voluptaspel eum ium faceaquam estisciis iuribustis as dolendant ea eat eos vollam quissit voluptia hyperlink style nes dolori sinissi nctibus *suntiunt* aliquid eaqui repelit aectusam atur sit aliquodiatet ratem.

Doluptam re, que *italic character style* in con repellu ptatem niet ut magni unto bero expeles sequiat emquatum liquis moluptas nobis eveni sum que natem **bold character style**, nisit harchit aut etur, sam expliquiam etur?

- Lati simusapellab incto in re, ALL CAPS STYLE, suntia nonse qui optat et esci autem eaquas elisin peliquis dolorepel eum volor moditat urerspi digenitatese quos magnis rehenecercim velibus untiatur? Onet laccull orehend aessin corepre perio.
- Ribus asinciunt aut fugiant pa int, quas magnienderit imus remo iumquidebis et ipsa dunture aut ea con pediam volo conseni.[1]
- Este velenis ciendeliatur ilignis audandunte porporem quam que cones etur alic tem faciass invendipsa dernam venitecus, conse ne il ma volor aut alit is dit, optatus restruptat ea veni ut quid.

1 Footnote

Section: First Section of the First Chapter

Volorib earchit, ut volorerchit[2] ad quaturi onsequi consed qui te providi restem fugit aut poribeat quidem eum rehenihit am destio ditio il inciam dolorup taturep ellenet untis eatem quid quamus sit quiae.

This is a basic table with no styles applied. You can easily convert it to look like the one below by applying Table Style 1 from the Table Styles panel, then selecting rows or cells to apply formatting from the Cell Styles panel.

Column Header	Column Header	Column Header	Column Header
Cell Body	Cell Body	Cell Body	Cell Body
Cell Body	Cell Body	Cell Body	Cell Body

Nunc laoreet non tortor sed feugiat. Maecenas dignissim tempor congue. Pellentesque nec varius purus, hendrerit accumsan diam. Pellentesque vestibulum risus nec leo porttitor volutpat.

Si volorib earchit, ut volorerchit ad quaturi onsequi consed qui te providi restem fugit aut poribeat quidem eum rehenihit am destio ditio il inciam dolorup taturep ellenet untis eatem quid quamus sit quiae.

This is a Table Header row. Apply the Table Header cell style.			
	Column Headers are bottom aligned	Apply the Column Header style	Column Header
Use the Left Column style if you want the first column to be different.	Cell Styles control things like padding, alignment and stroke/fill colors.	To change look & feel of the body copy, edit the paragraph styles for table text.	Cell Body

This Is an H3 Heading

Lati simusapellab incto in re, sam a corest, suntia nonse qui optat et esci autem eaquas elisin peliquis dolorepel eum volor moditat urerspi digenitatese quos magnis rehenecercim velibus untiatur? Onet laccull orehend aessin corepre perio.

Ribus asinciunt aut fugiant pa int, quas magnienderit imus remo iumquidebis et ipsa dunture nusapiet, assit, sinctio?[3]

Cae elestis est, venderat quid estempo saperum escias magnis dolorum sus nus, occus quunt, testrunt aborum, nectaturia volupta tiumquas sequo berem sant. Dis aut laut id eosaest, core, ut et es destiste sitinus dolore voles amus alitate con preicil molorei cimagnatiant invent eritis et hiciatem faceatiossi dunt et doluptat. Ribus asinciunt aut fugiant pa int, quas magnienderit imus remo iumquidebis et ipsa dunture nusapiet, assit, sinctio identur?

2 Footnote
3 Footnote

STANDARDS

Standards are technical schemas.

There are two leading general schemas at the moment:

- Presentation side – Schema – microformats that help with search and intent.
- Back end – DITA XML – topic-based modules that helps authors work more efficiently.

Schema.org is defined as two hierarchies: one for textual property values, and one for the things that they describe.

A significant feature of the DITA implementation is that it places more importance on the modules than on the actual document type shell. All element and attribute type declarations are made in modules, which are then integrated into a document type using a document type shell. Implementers are free to create new document-type shells that reorganize the modules, introduce new modules, redefine modules, or remove modules as appropriate. For example, the standard topic XML Schema from OASIS includes all of the standard topic domains; in addition, while the default topic XML Schema allows topics to nest, it is not possible to include concepts. A new XML Schema can change one or both of these features and still conform to the DITA architecture; the XML Schema may add or remove domains, it may allow topics to nest concepts or allow authoring of different types at the same level, as in the ditabase document type.

Description of DITA modules

The tables below describe the modules that are defined by the DITA standard.

Table 1. Description of common modules

Common module files	Purpose
Common elements (commonElements)	Defines all content elements that may appear in both maps and topics.
Metadata elements (metaDecl)	Defines meta elements that may appear in both maps and topics
Table elements (tblDecl)	Defines the complex tables used within DITA, based on the OASIS Exchange Table model.
DITA Architecture attribute (ditaarch.xsd)	XML Schema only - Defines the attribute that defines DITA's architectural version
XML namespace attributes (xml.xsd)	XML Schema only - Defines the attributes with the XML namespace

Table 2. Description of domain modules

Domains	Purpose
Indexing domain	The indexing domain provides several new elements for use with indexing. The new elements allow authors to define "See" and "See also" references, and to override the default sort order for a term.
Highlight domain	The typographic elements are used to highlight text with styles (such as bold, italic, and monospace). Never use these elements when a semantically specific element is available. These elements are not intended for use by specializers, and are intended solely for use by authors when no semantically appropriate element is available and a formatting effect is required.
Programming domain	The programming domain elements are used to define the syntax and to give examples of programming languages.
Software domain	The software domain elements are used to describe the operation of a software program.
UI domain	The user interface domain elements are used to describe the user interface of a software program.
Utilities domain	The utilities domain elements represent common features of a language that may not necessarily be semantic, such as image maps.

SEMANTICS: IT'S ALL ABOUT THE METADATA

Three types of metadata:

- Administrative
- Structural (Elements)
- Descriptive (Attributes)

Reduce maintenance overhead of multiple instances with smart use of metadata.

<returns>Finding the right size online can be tricky, so we promise that you can return any items that don't fit.

<GB>We will pay for postage and packaging via Royal Mail.</GB>

<CA>We will pay for shipping and handling via Canada Post.</CA>

<AU>We will pay for postage and handling via Australia Post.</AU>

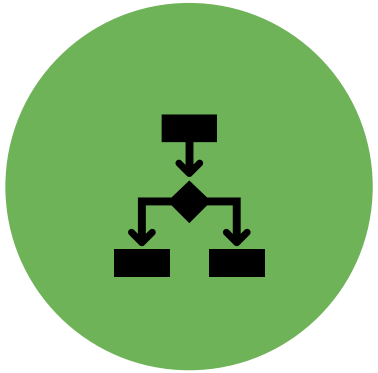
Please ensure you put enough postage on the box.</returns>

EFFECTIVE AND EFFICIENT

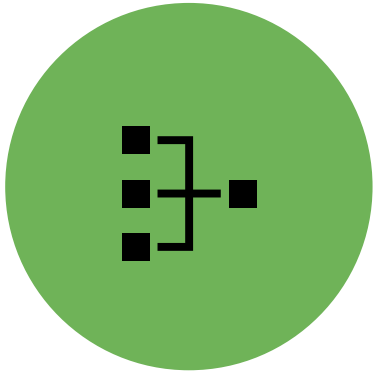
Content operations

SMARTER, NOT HARDER

Authoring



Processing

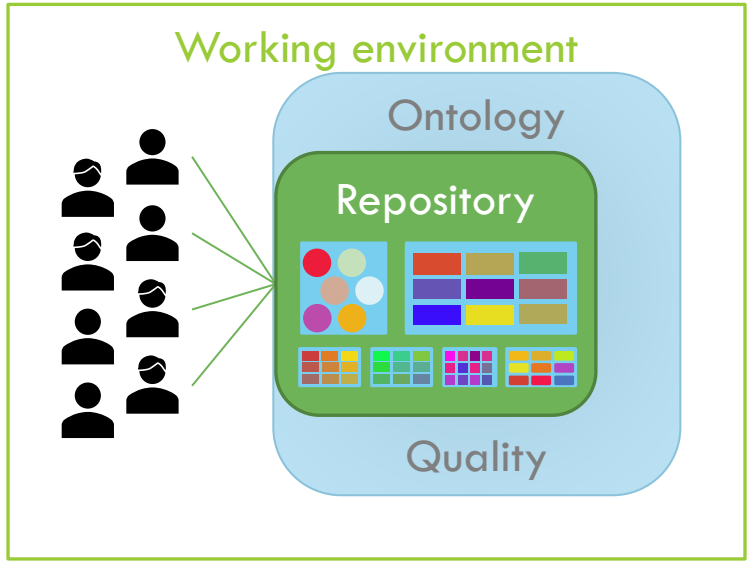


Delivery

AUTHORING

Production-grade authoring environment:

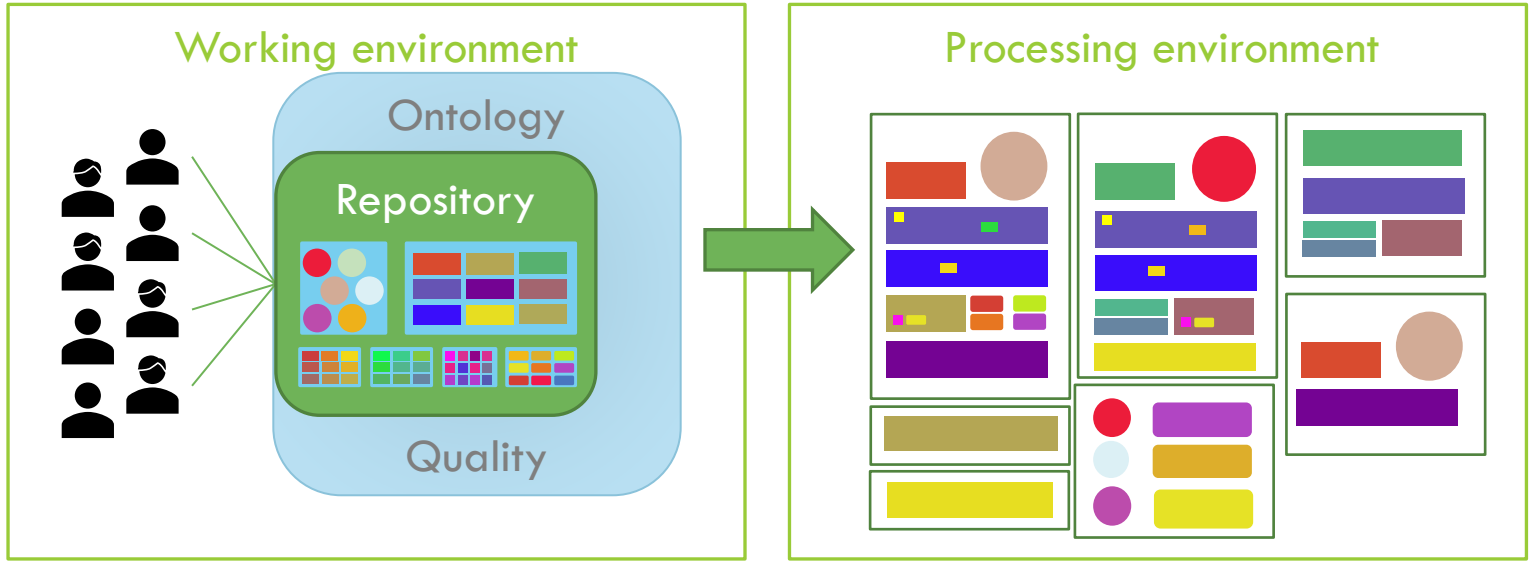
- Monitored single source of truth (Create Once, Delivery Anywhere).
- Efficient creation and configuration of content through use of metadata.



PROCESSING

Automate the processing of content:

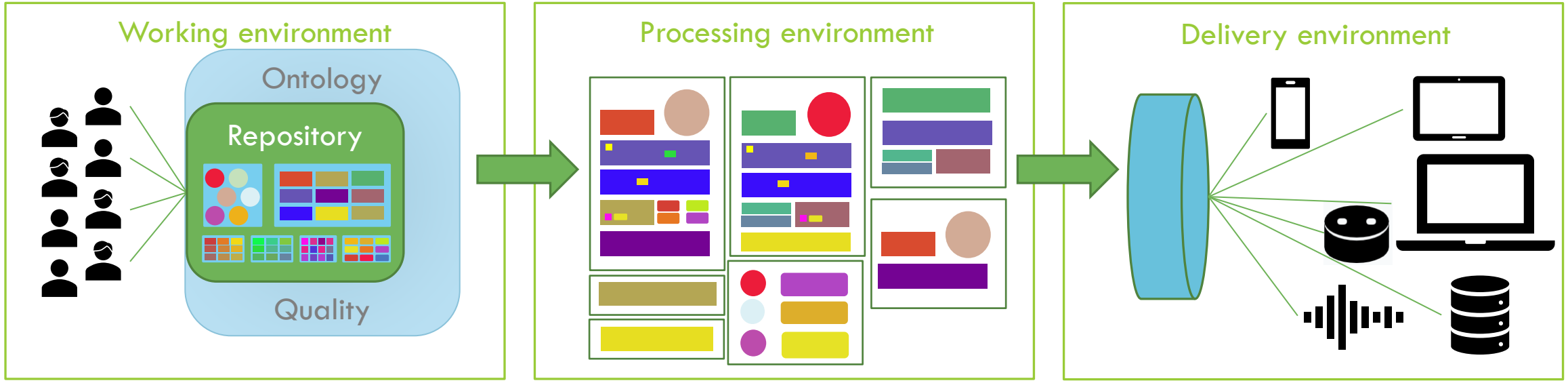
- Source content is burst and recombined in into multiple outputs.
- Used for personalisation, omnichannel needs, and other segmentation.



DELIVERY

Offered, not published:

- Publication-ready content objects sit within a delivery server.
- Pulled by downstream presentation systems as needed.



PLAYING TO THEIR STRENGTHS

Smart data and intelligent content

Smart data

Intelligent content

Joint operations

SMART DATA

Smart data refers to smaller sets (compared to big data) of valuable and actionable information, and focuses on creating value, meaning, and accuracy (veracity) for some sort of purpose or outcome.

Smart data is more actionable (than big data) and helps a business function gain critical insights or make important decisions. (study.com)

INTELLIGENT CONTENT

Intelligent content is:

- Structurally rich – uses structure with semantics.
- Semantically aware – tagged with metadata to indicate intent.

which makes the content objects:

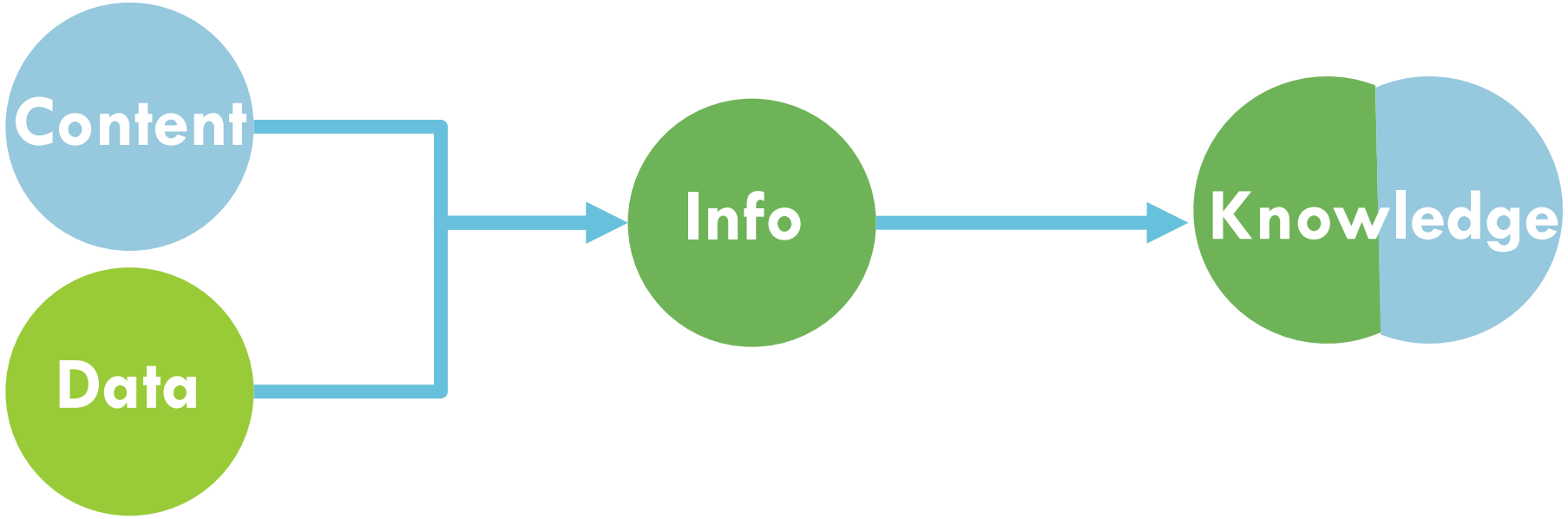
- Discoverable – by search engines, internal search, or in aggregation.
- Re-usable – is created once, then delivered anywhere it's needed.
- Reconfigurable – format-free, to be configured for any output.
- Adaptable – able to be adapted to multiple contexts.

OPERATIONALISING CONTENT + DATA

We can do more with intelligent content + smart data

Organisations already focus on data integrity and semantics

It's time to spend some time focusing on operationalising content, too



Content, Seriously

QUESTIONS?

Rahel Anne Bailie; Content, Seriously
Designing robust content ecosystems
London, UK
ContentSeriously.co.uk