



Making of gold

Stories from the content enrichment
workbench



Objectives of this presentation

- Explore the content enrichment challenge
 - Motivation
 - Typical problems (and some solutions)
 - Case studies
 - Lessons learnt



Motivation for Content Enrichment

Changes in customer expectations

Challenges to existing business models

New revenue opportunities

Maintain relevance as a publisher



What is happening in this area?

Elsevier's Article of the Future

OpenCalais

data.gov.uk

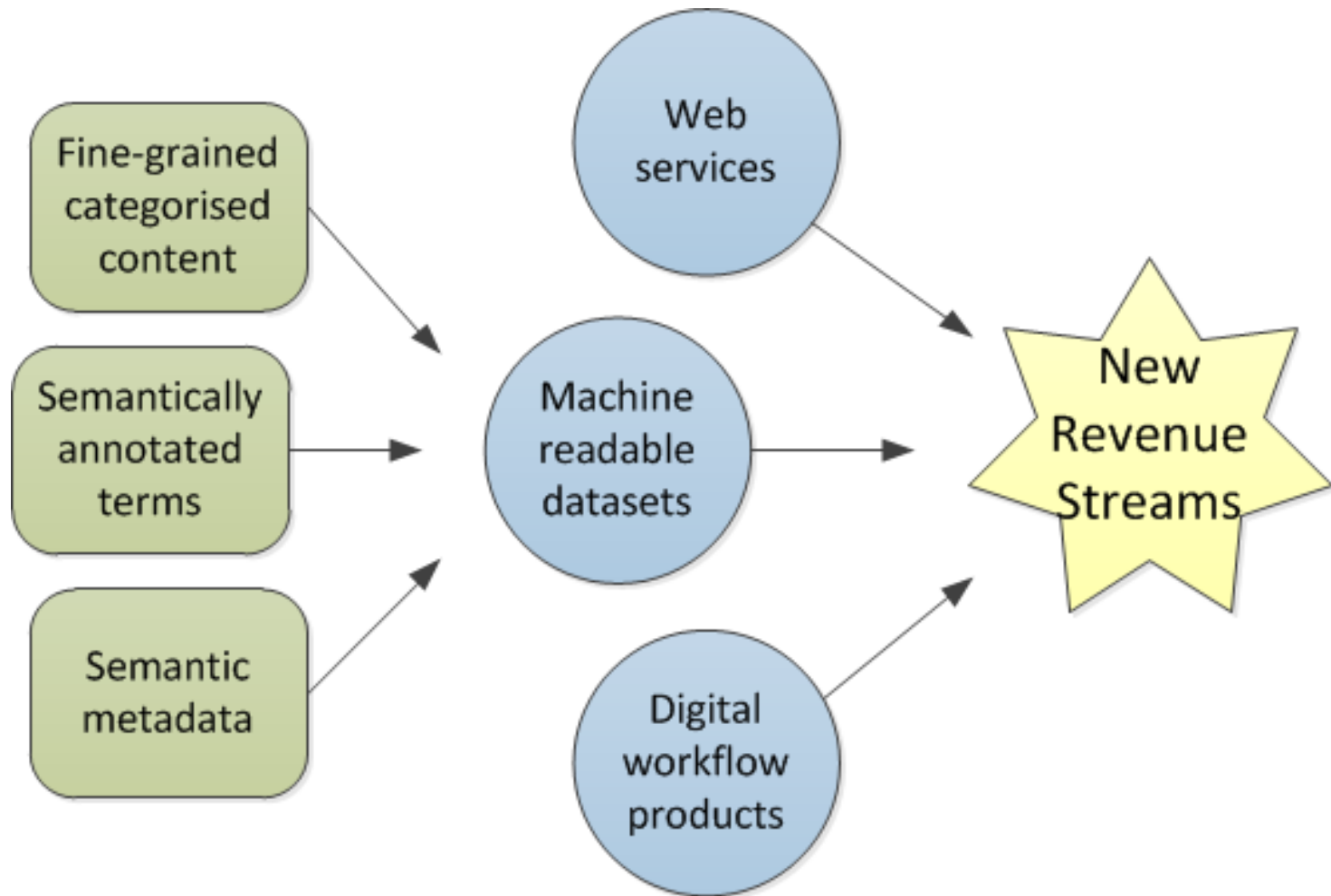
The wider web

What are the typical challenges?

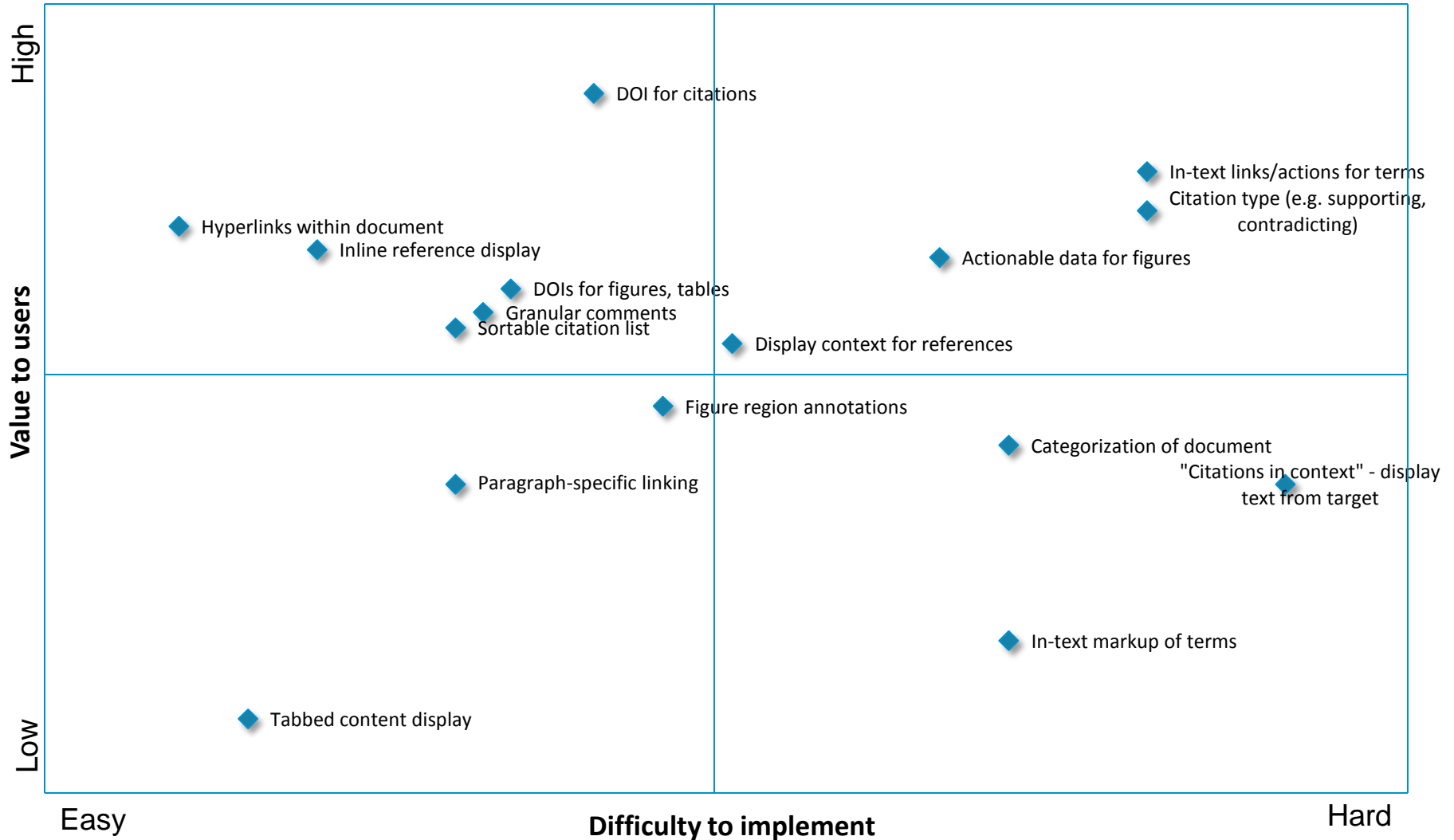
- What balance of automated and manual processes are required?
- What should we enrich?
- What will increase the usage and value of our content?
- Is there a business model for enrichment?
- How and where do we do enrichment?
- How can we take advantage of third-party user enhancements?
- How do we need to rework our publication process to support enrichment?



What to enrich?



What to enrich?

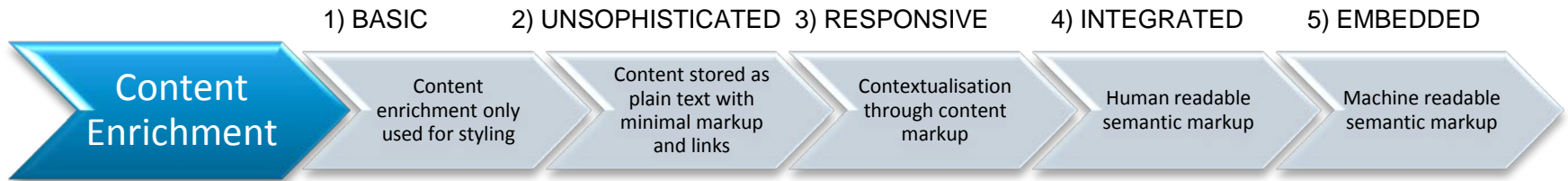


Methods of content enrichment

1. Automated recognition of terms and structures
2. Manual annotation
3. Retain the authors' original rich data
4. Outsourced enrichment
5. User generated enrichment



Content enrichment maturity



- Content enrichment needs to be considered a core organisational capability
- Need to determine the minimum level of maturity you need to reach to achieve your organisational objectives
- Need to start learning now as this will require significant organisational change that will take time

Case studies

- Pharmaceutical Press
 - Supporting manual enrichment of content
- BSI
 - Managing outsourced content enrichment
 - Automatic conversion of British Standards



Supporting manual enrichment of content

BNF Preparation - Mozilla Firefox

BNF Preparation

Aspirin

Status: Valid
Last saved: 09:51:28

Actions

- Add presentation
- Save and continue editing
- Save and close
- Cancel checkout and close without saving
- Show preview

Preparation details

Preparation Name: Aspirin

Sub-title:

Index text: IX ✕

Add index-text

Add additional text

Manufacturers: ✕

Add manufacturer

DM+D Status: 😊 VTM ID is set

Legal Status:

NHS Status:

Type:

Black triangle status? | Less suitable flag?

Tablets | Tablets | **Dispersible tablets** | Suppositories

Presentation

Presentation type:


Manufacturer's type:

Presentation flags

<input type="checkbox"/> Modified release?	<input type="checkbox"/> Scored?	<input type="checkbox"/> Chewable?
<input type="checkbox"/> Sugar free?	<input type="checkbox"/> Gluten free?	<input type="checkbox"/> Black triangle?
<input type="checkbox"/> Less suitable?	<input type="checkbox"/> Not on NHS?	

Presentations

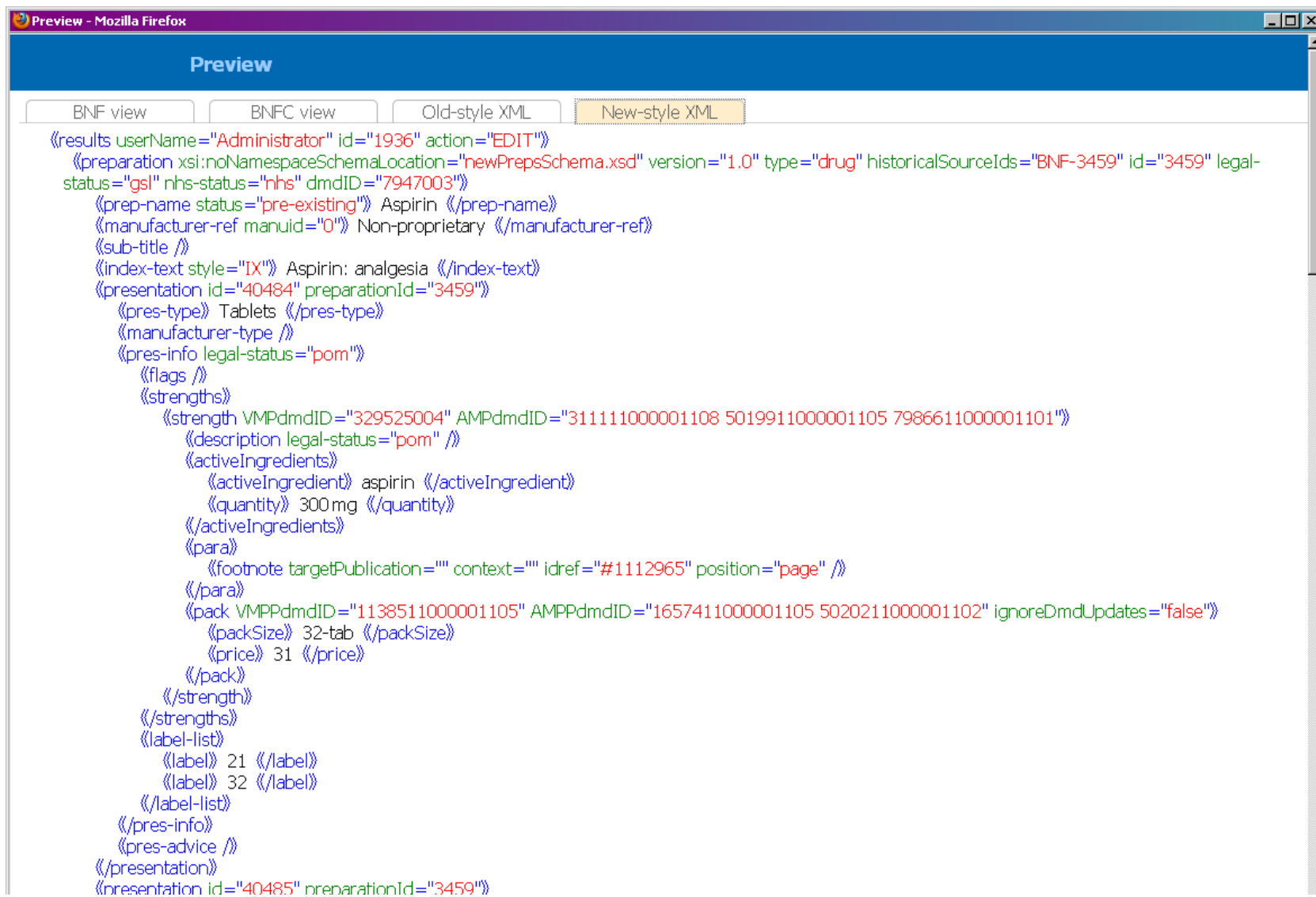
- Tablets
- Tablets
- Dispersible tablets**
- Suppositories



Supporting manual enrichment of content

```
«preparation xsi:noNamespaceSchemaLocation="tightenedExistingPrepsSchema.xsd" id="3459" legal-status="gsl" nhs-status="nhs"»
  «prep-name status="pre-existing"» Aspirin «/prep-name»
  «manufacturer-ref manuid="0"» Non-proprietary «/manufacturer-ref»
  «index-text style="IX"» Aspirin: analgesia «/index-text»
  «presentation»
    «pres-type» Tablets «/pres-type»
    «pres-info»
      «PoM /»
    , aspirin 300 mg
      «footnote targetPublication="" context="" idref="#1112965" position="page" /»
    , net price 32-tab = 31 p.
      «label-list»
        «label» 21 «/label»
        «label» 32 «/label»
      «/label-list»
    «/pres-info»
  «/presentation»
  «presentation»
    «pres-type» Tablets «/pres-type»
    «pres-info»
      «PoM /»
    , e/c, aspirin 300 mg
      «footnote targetPublication="" context="" idref="#1112965" position="page" /»
    , net price 100-tab = £4.83.
      «label-list»
        «label» 5 «/label»
        «label» 25 «/label»
        «label» 32 «/label»
      «/label-list»
    «/pres-info»
  «/presentation»
  «presentation»
    «pres-type» Dispersible tablets «/pres-type»
    «pres-info»
      «PoM /»
    aspirin 300 mg
```

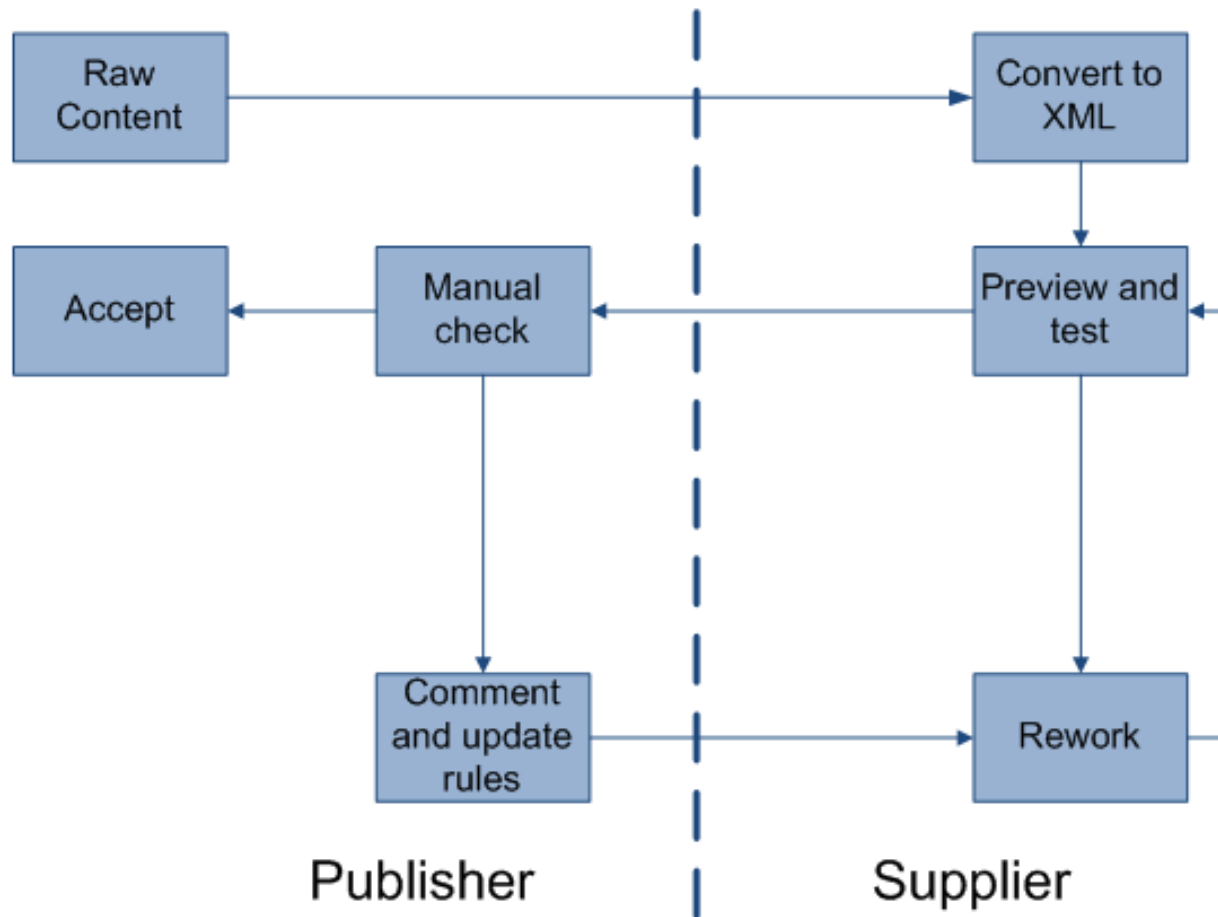
Supporting manual enrichment of content



The screenshot shows a Mozilla Firefox browser window titled "Preview - Mozilla Firefox". The browser displays an XML document in "New-style XML" format. The XML content is a structured representation of a drug record, including details like user information, preparation details, manufacturer information, and active ingredients.

```
«results userName="Administrator" id="1936" action="EDIT"»
  «preparation xsi:noNamespaceSchemaLocation="newPrepsSchema.xsd" version="1.0" type="drug" historicalSourceIds="BNF-3459" id="3459" legal-
    status="gsl" nhs-status="nhs" dmdID="7947003"»
    «prep-name status="pre-existing"» Aspirin «/prep-name»
    «manufacturer-ref manuid="0"» Non-proprietary «/manufacturer-ref»
    «sub-title /»
    «index-text style="IX"» Aspirin: analgesia «/index-text»
    «presentation id="40484" preparationId="3459"»
      «pres-type» Tablets «/pres-type»
      «manufacturer-type /»
      «pres-info legal-status="pom"»
        «flags /»
        «strengths»
          «strength VMPdmdID="329525004" AMPdmdID="311111000001108 5019911000001105 7986611000001101"»
            «description legal-status="pom" /»
            «activeIngredients»
              «activeIngredient» aspirin «/activeIngredient»
              «quantity» 300 mg «/quantity»
            «/activeIngredients»
            «para»
              «footnote targetPublication="" context="" idref="#1112965" position="page" /»
            «/para»
            «pack VMPPdmdID="1138511000001105" AMPPdmdID="1657411000001105 5020211000001102" ignoreDmdUpdates="false"»
              «packSize» 32-tab «/packSize»
              «price» 31 «/price»
            «/pack»
          «/strength»
        «/strengths»
        «label-list»
          «label» 21 «/label»
          «label» 32 «/label»
        «/label-list»
      «/pres-info»
      «pres-advice /»
    «/presentation»
  «presentation id="40485" preparationId="3459"»
```

Managing outsourced content enrichment



Managing outsourced content enrichment

Sentinel

HTML Preview XML

View Previous HTML

Table of contents

[UK National Annex to Eurocode 4: Design of composite steel and concrete structures](#)
[National Annex \(informative\) to BS EN 1994-1-1:2004, Eurocode 4: Design of composite steel and concrete structures — Part 1.1: General rules and rules for buildings](#)

[Introduction](#)

[NA.1 Scope](#)

[NA.2 Nationally Determined Parameters](#)

[NA.2.1 General](#)

[NA.2.2 Partial factor, \$\gamma_p\$ \[BS EN 1994-1-1:2004, 2.4.1.1\(1\)\]](#)

[NA.2.3 Partial factor, \$\gamma_v\$ \[BS EN 1994-1-1:2004, 2.4.1.2\(5\)\]](#)

[NA.2.4 Partial factor, \$\gamma_{vs}\$ \[BS EN 1994-1-1:2004, 2.4.1.2\(6\)\]](#)

[NA.2.5 Partial factor, \$\gamma_{Mf,s}\$ \[BS EN 1994-1-1:2004, 2.4.1.2\(7\)\]](#)

[NA.2.6 Shrinkage of concrete \[BS EN 1994-1-1:2004, 3.1\(4\)\]](#)

[NA.2.7 Nominal thickness, \$t\$, of steel sheets \[BS EN 1994-1-1:2004, 3.5\(2\)\]](#)

[NA.2.8 Other types of steel section \[BS EN 1994-1-1:2004, 6.4.3\(1\)\(h\)\]](#)

[NA.2.9 Partial factor, \$\gamma_v\$ \[BS EN 1994-1-1:2004, 6.6.3.1\(1\)\]](#)

[NA.2.10 Splitting forces in the direction of the slab thickness \[BS EN 1994-1-1:2004, 6.6.3.1\(3\)\]](#)

[NA.2.11 Means of anchorage for non-continuous sheeting \[BS EN 1994-1-1:2004, 6.6.4.1\(3\)\]](#)

[NA.2.12 Partial factor, \$\gamma_{Mf,s}\$ \[BS EN 1994-1-1:2004, 6.8.2\(1\)\]](#)

[NA.2.13 Partial factor, \$\gamma_{Ff}\$ \[BS EN 1994-1-1:2004, 6.8.2\(2\)\]](#)

[NA.2.14 Limit for narrowly spaced webs, \$b/r/s\$ \[BS EN 1994-1-1:2004, 9.1.1\(2\)\]](#)

[NA.2.15 Limit for the deflection of sheeting, \$\delta_{s,max}\$ \[BS EN 1994-1-1:2004, 9.6\(2\)\]](#)

[NA.2.16 Partial factor, \$\gamma_{Vs}\$ \[BS EN 1994-1-1:2004, 9.7.3\(4\)\]](#)

[NA.2.17 Partial factor, \$\gamma_{Vs}\$ \[BS EN 1994-1-1:2004, 9.7.3\(8\)\]](#)

[NA.2.18 Nominal factor, \$\mu\$ \[BS EN 1994-1-1:2004, 9.7.3\(9\)\]](#)

[NA.2.19 Partial factor, \$\gamma_v\$ \[BS EN 1994-1-1:2004, B.2.5\(1\)\]](#)

[NA.2.20 Partial factor, \$\gamma_{Vs}\$ \[BS EN 1994-1-1:2004, B.3.6\(5\)\]](#)

[NA.3 Decisions on the status of informative annexes](#)

[NA.4 References to non-contradictory complementary information](#)

[Bibliography](#)

The Emperor Penguin (*Aptenodytes forsteri*) is the tallest and heaviest of all living penguin species and is endemic to Antarctica. The male and female are similar in plumage and size, reaching 122 cm (48 in) in height and weighing anywhere from 22 to 45 kg (49 to 99 lb). The dorsal side and head are black and sharply delineated from the white belly, pale-yellow breast and bright-yellow ear patches. Like all penguins it is flightless, with a streamlined body, and wings stiffened and flattened into flippers for a marine habitat.

Its diet consists primarily of fish, but can also include crustaceans, such as krill, and cephalopods, such as squid. In hunting, the species can remain submerged up to 18 minutes, diving to a depth of 535 m (1,755 ft). It has several adaptations to facilitate this, including an unusually structured hemoglobin to allow it to function at low oxygen levels, solid bones to reduce barotrauma, and the ability to reduce its metabolism and shut down non-essential organ functions.

The Emperor Penguin is perhaps best known for the sequence of journeys adults make each year in order to mate and to feed their offspring. The only penguin species that breeds during the Antarctic winter, it treks 50–120 km (31–75 mi) over the ice to breeding colonies which may include thousands of individuals. The female lays a single egg, which is incubated by the male while the female returns to the sea to feed; parents subsequently take turns foraging at sea and caring for their chick in the colony. The lifespan is typically 20 years in the wild, although observations suggest that some individuals may live to 50 years of age.

Automatic conversion of British Standards

EN 1991-1-5_2003.doc [Compatibility Mode] - Microsoft Word

Home Insert Page Layout References Mailings Review View MathType

Paste Clipboard

Arial 12

Font Paragraph Styles Editing

Find Replace Select

Change Styles

1.5.5
initial temperature T_0
the temperature of a structural element at the relevant stage of its restraint (completion)

1.5.6
cladding
the part of the building which provides a weatherproof membrane. Generally cladding will only carry self weight and/or wind actions

1.5.7
uniform temperature component
the temperature, constant over the cross section, which governs the expansion or contraction of an element or structure (for bridges this is often defined as the "effective" temperature, but the term "uniform" has been adopted in this part)

1.5.8
temperature difference component
the part of a temperature profile in a structural element representing the temperature difference between the outer face of the element and any in-depth point

1.6 Symbols

(1) For the purposes of this Part of Eurocode 1, the following symbols apply.

NOTE: The notation used is based on ISO 3898

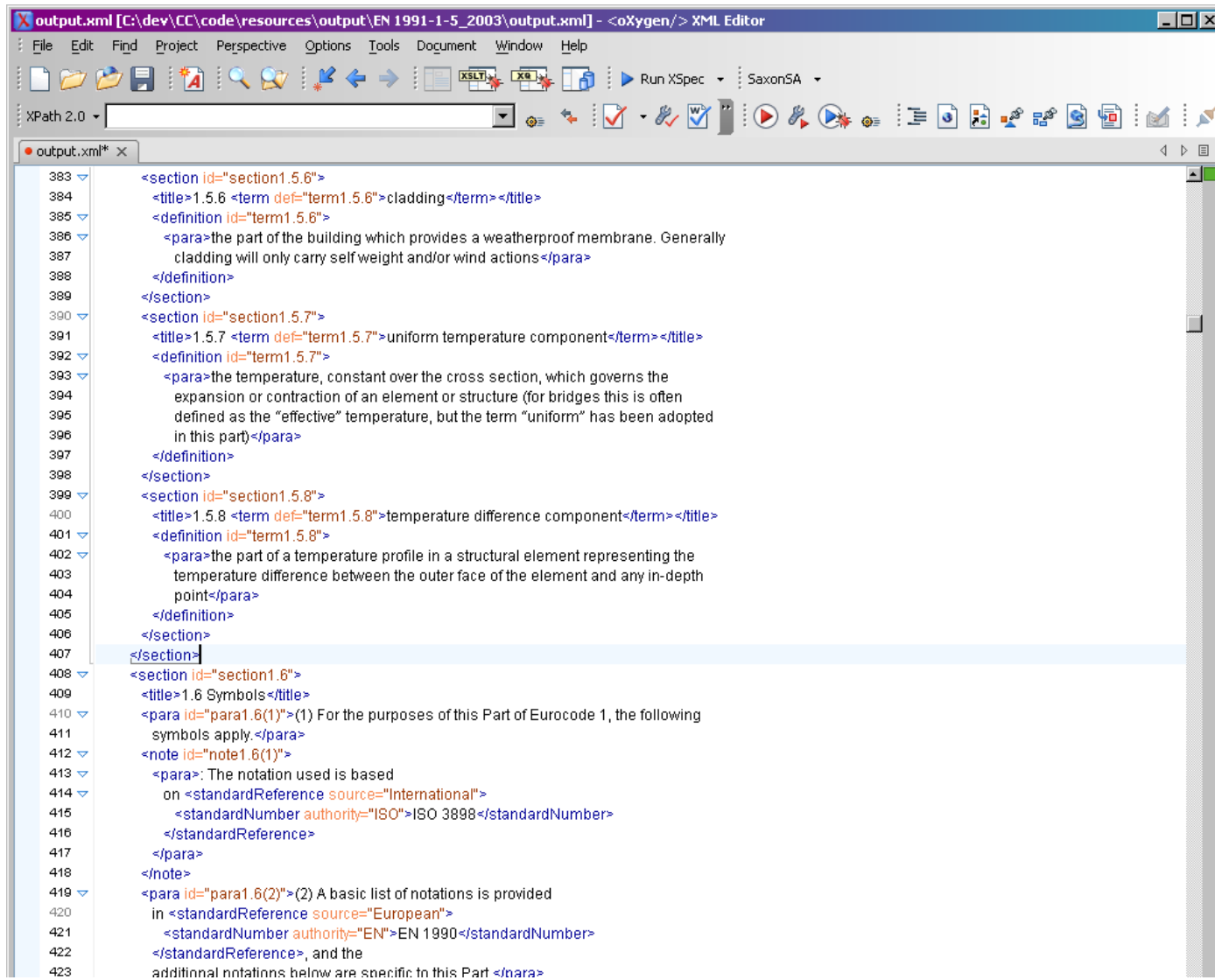
(2) A basic list of notations is provided in EN 1990, and the additional notations below are specific to this Part.

Latin upper case letters

R thermal resistance of structural element

R_i thermal resistance at the inner surface

Automatic conversion of British Standards

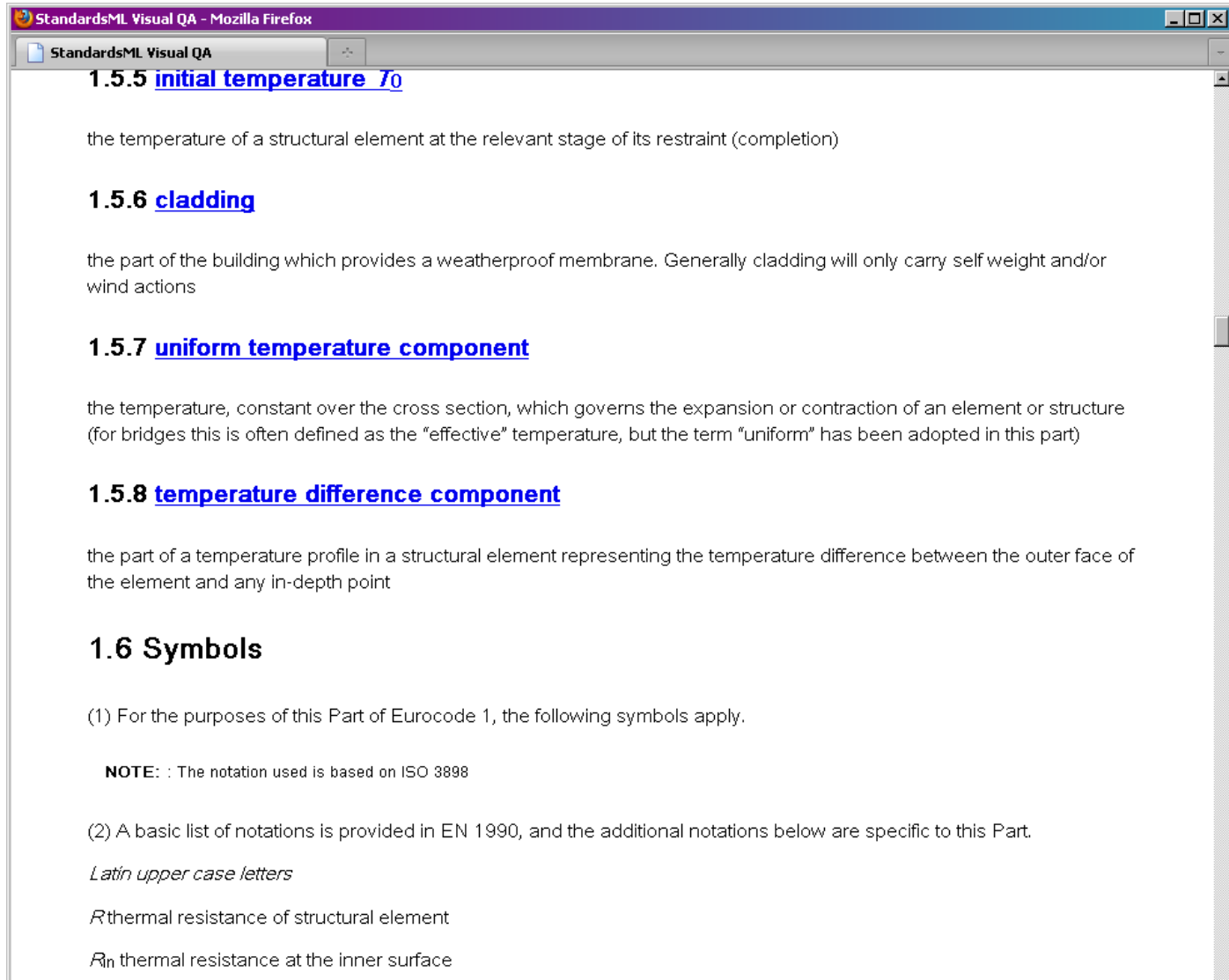


The screenshot displays the oXygen XML Editor interface. The title bar indicates the file is 'output.xml' located at 'C:\dev\CC\code\resources\output\EN 1991-1-5_2003\output.xml', opened with the '<oXygen/> XML Editor'. The menu bar includes File, Edit, Find, Project, Perspective, Options, Tools, Document, Window, and Help. The toolbar contains various icons for file operations, navigation, and execution. The main editing area shows the XML content of 'output.xml' with line numbers on the left. The XML structure includes sections for 'cladding' (1.5.6), 'uniform temperature component' (1.5.7), 'temperature difference component' (1.5.8), and 'Symbols' (1.6). The 'Symbols' section contains two paragraphs: paragraph 1.6(1) regarding the notation used based on ISO 3898, and paragraph 1.6(2) regarding a basic list of notations provided in EN 1990. The XML uses standard tags like <section>, <title>, <definition>, <para>, <note>, and <standardReference> to structure the content.

```
383 <section id="section1.5.6">
384 <title>1.5.6 <term def="term1.5.6">cladding</term></title>
385 <definition id="term1.5.6">
386 <para>the part of the building which provides a weatherproof membrane. Generally
387 cladding will only carry self weight and/or wind actions</para>
388 </definition>
389 </section>
390 <section id="section1.5.7">
391 <title>1.5.7 <term def="term1.5.7">uniform temperature component</term></title>
392 <definition id="term1.5.7">
393 <para>the temperature, constant over the cross section, which governs the
394 expansion or contraction of an element or structure (for bridges this is often
395 defined as the "effective" temperature, but the term "uniform" has been adopted
396 in this part)</para>
397 </definition>
398 </section>
399 <section id="section1.5.8">
400 <title>1.5.8 <term def="term1.5.8">temperature difference component</term></title>
401 <definition id="term1.5.8">
402 <para>the part of a temperature profile in a structural element representing the
403 temperature difference between the outer face of the element and any in-depth
404 point</para>
405 </definition>
406 </section>
407 </section>
408 <section id="section1.6">
409 <title>1.6 Symbols</title>
410 <para id="para1.6(1)">(1) For the purposes of this Part of Eurocode 1, the following
411 symbols apply.</para>
412 <note id="note1.6(1)">
413 <para>: The notation used is based
414 on <standardReference source="International">
415 <standardNumber authority="ISO">ISO 3898</standardNumber>
416 </standardReference>
417 </para>
418 </note>
419 <para id="para1.6(2)">(2) A basic list of notations is provided
420 in <standardReference source="European">
421 <standardNumber authority="EN">EN 1990</standardNumber>
422 </standardReference>, and the
423 additional notations below are specific to this Part </para>
```



Automatic conversion of British Standards



StandardsML Visual QA - Mozilla Firefox

StandardsML Visual QA

1.5.5 initial temperature T_0

the temperature of a structural element at the relevant stage of its restraint (completion)

1.5.6 cladding

the part of the building which provides a weatherproof membrane. Generally cladding will only carry self weight and/or wind actions

1.5.7 uniform temperature component

the temperature, constant over the cross section, which governs the expansion or contraction of an element or structure (for bridges this is often defined as the "effective" temperature, but the term "uniform" has been adopted in this part)

1.5.8 temperature difference component

the part of a temperature profile in a structural element representing the temperature difference between the outer face of the element and any in-depth point

1.6 Symbols

(1) For the purposes of this Part of Eurocode 1, the following symbols apply.

NOTE: : The notation used is based on ISO 3898

(2) A basic list of notations is provided in EN 1990, and the additional notations below are specific to this Part.

Latin upper case letters

R thermal resistance of structural element

R_n thermal resistance at the inner surface

Lessons learnt / our approach to cost-effective enrichment

- Derive enrichment targets from organisational objectives and desired content capabilities
- Reduce the cost of manual enrichment with tools that limit options, provide validation, and provide preview
- Manage outsourced suppliers with automation and manual checks, and increase the automation over time
 - Expose content checks to suppliers so they can test themselves
- Use automated enrichment where possible
 - And supplement with automatic testing to focus manual checks – it's much easier to detect problems than fix them
- Start learning now as this is an organisation behaviour and capability change





67 Bricks Ltd.
<http://www.67bricks.com/>

Sam Herbert
sam.herbert@67bricks.com
+44 (0)7734 138274