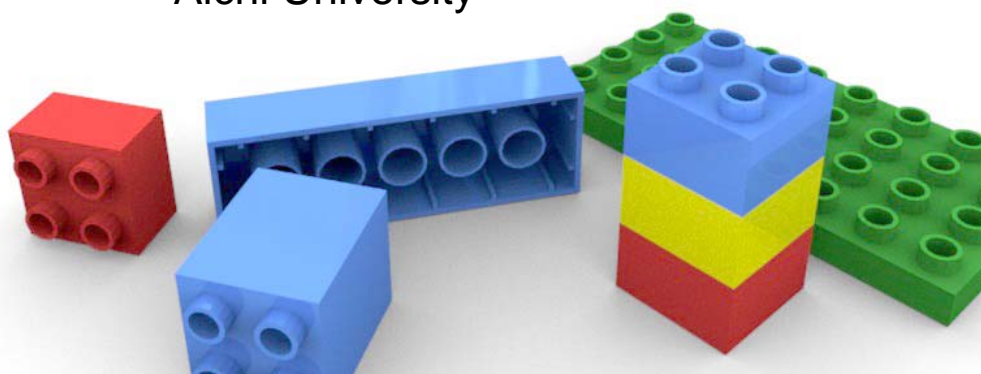


# Implementing XML for Japanese-Language Scholarly Articles

2012. 10. 16

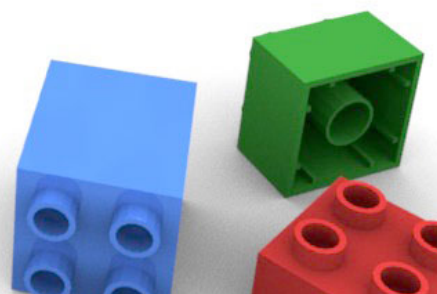
Soichi Tokizane

Aichi University



## Whom Do I Represent?

- Scholarly XML Publishing Association
  - Newly-formed organization to promote XML for scholarly publishing in Japan

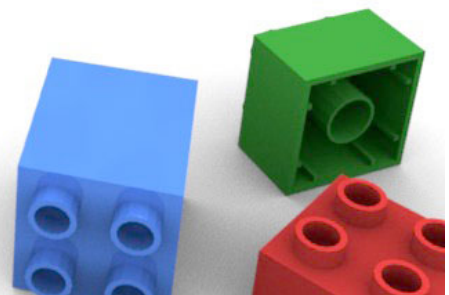


# SGML/XML for Scholarly Publishing in Japan

- There were attempts to implement SGML to scholarly publishing in early 90's.
- XML was tried, too, in early 2000's

3

JATS-Con 2012

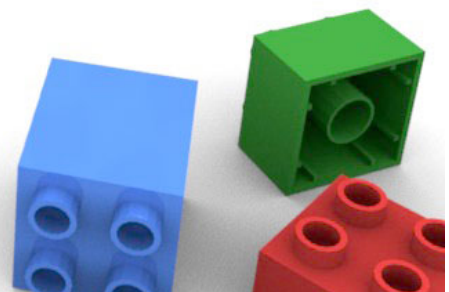


## J-STAGE

- J-STAGE, an ejournal platform of JST
  - Launched in 1999
  - Developed J-STAGE SGML DTD
    - Based on JICST-DTD
  - FrameMaker+SGML production workflow
  - Investigated XML in 2000, but found it too early to implement it

4

JATS-Con 2012

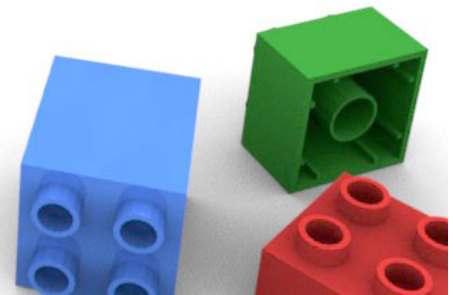


# JATS 0.4 for New J-STAGE

- New version of J-STAGE launched in May, 2012
  - Implemented JATS 0.4

5

JATS-Con 2012



But all three languages are different  
from each other

- Japanese language came from Mongolia
- Japanese language owes
  - characters to China
  - old words to Korea



6

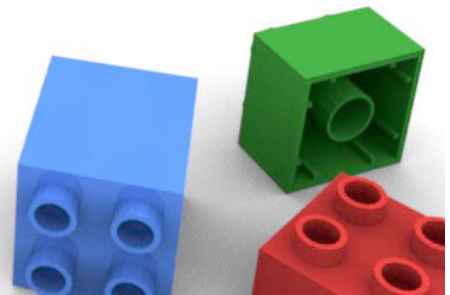


## In any case ...

- We, three, all needed multi-language extension of NLM DTD

7

JATS-Con 2012

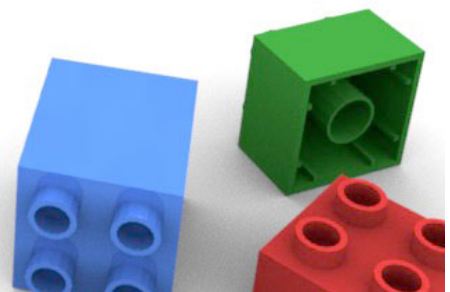


## Characteristics of Japanese Writing (1)

- Scripts
  - 野田 佳彦 (Kanji – Chinese)
  - のだ よしひこ (Hirakana)
  - ノダ ヨシヒコ (Katakana)
  - Yoshihiko Noda (Romaji – Roman)

8

JATS-Con 2012

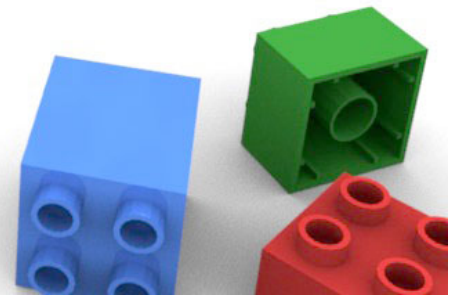


## Characteristics of Japanese Writing (2)

- Name Order
  - 野田 佳彦 (Kanji – Chinese)
  - Yoshihiko Noda (Romaji – Roman)
    - (Family names in red)
- In JATS
  - name-style="eastern"

9

JATS-Con 2012

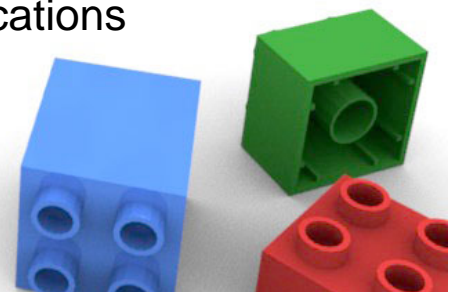


## Characteristics of Japanese Writing (3)

- Vertical, Right-to-Left
  - Traditional (also in historic Chinese)
  - Used in newspapers, magazines, books and humanity/social science publications/textbooks
- Western style
  - New
  - Exclusively used in business
  - Used in science/technology publications

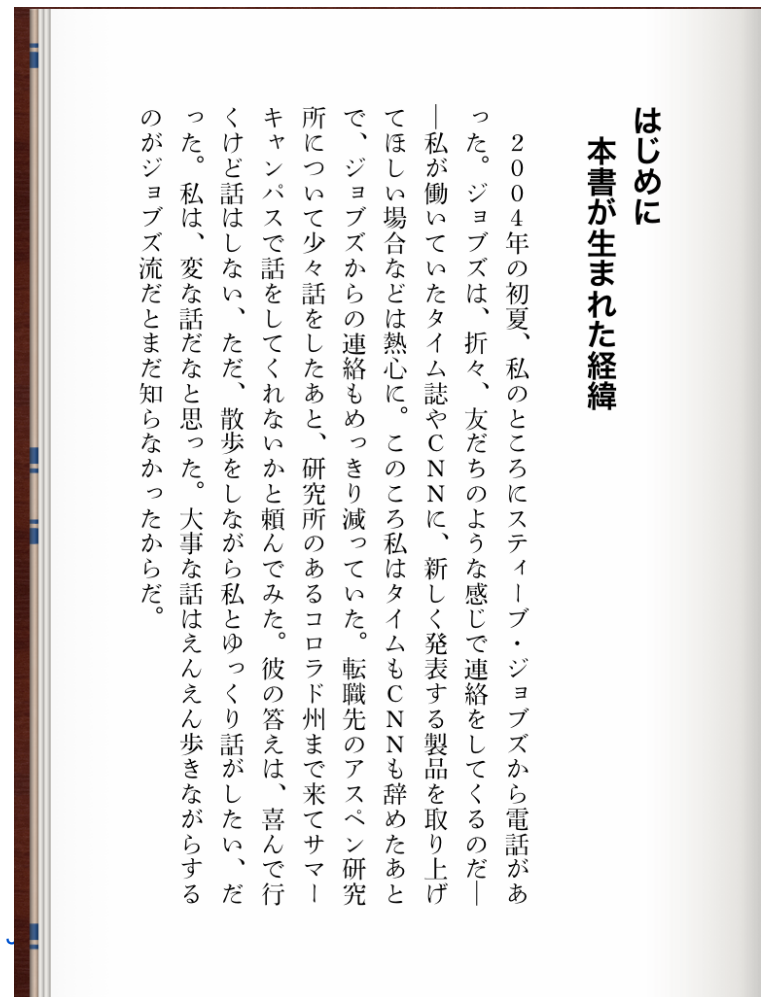
10

JATS-Con 2012



## Vertical, Right-to-Left

11



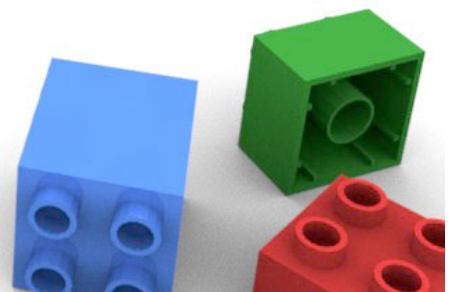
## Parentheses rotated in Vertical Writing

(あいうえお)

(かきくけこ)

12

JATS-Con 2012



# Ruby

- Used to express pronunciations
- Used to teach children how to read
- Needed, e.g., for personal names as well

- 時実 象一

- can be pronounced as, “Tokijitsu Zoichi”

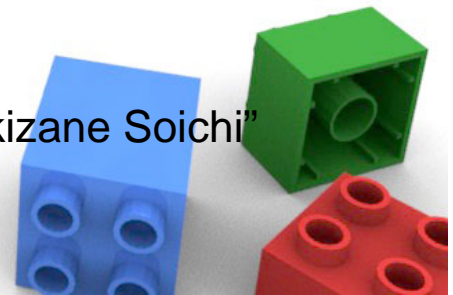
ときざね そういち

- 時実 象一

- so that it is read correctly as, “Tokizane Soichi”

13

JATS-Con 2012



3 坊っちゃん

## Ruby

- Used to express pronunciations
  - In smaller fonts

していたら、同級生の一人が冗談にじようだんいくらか威張いばっても、  
そこから飛び降りる事は出来まい。弱虫やーい。と囃はやした  
からである。小使こつかいに負ぶさって帰って来た時、おやじが大  
きな眼めをして二階ぐらいから飛び降りて腰を抜かす奴やつがあ  
るかと云いったから、この次は抜かさずに飛んで見せますと  
答えた。

親類かざのものから西洋製のナイフを貰もらって奇麗きれいな刃はを日に  
翳かざして、友達ともだちに見せていたら、一人が光る事は光るが切れ  
そうもないと云った。切れぬ事があるか、何でも切ってみ

14



## Emphasis

- *Italics* or **bold faces** do not exist in Japanese writings
- Emphasis are used instead

あいうえお

かきくけこ

15

JATS-Con 2012



## Warichu

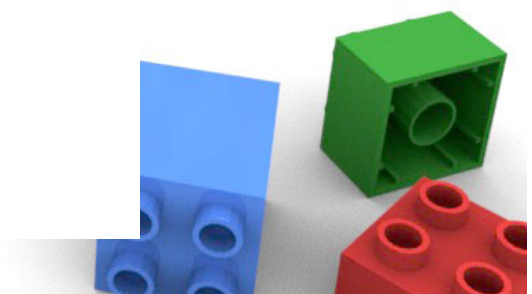
- Short note inserted within a sentence in two lines

かきくけこ (これは見  
本です)

か  
き  
く  
け  
こ  
(これは見  
本です)

16

JATS-Con 2012





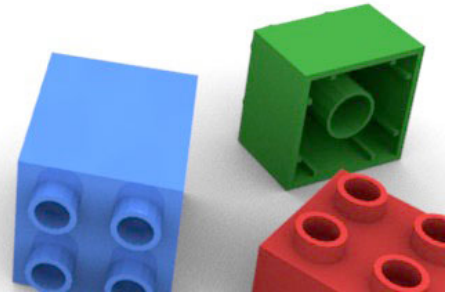
# Non-Gregorian Years

- Japanese traditional year designations
  - Now called, “Emperor year”, and is formal
- Rare in scholarly publications
  - Still appear in citations of old references

昭和四〇年 = 1965<sub>レ</sub>

17

JATS-Con 2012

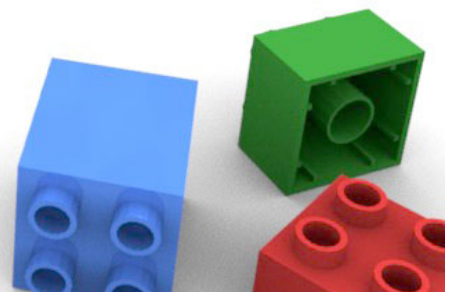


## Japanese STM Publishing (1)

- Number of journal titles (2008)
  - 3,047 total in STM
  - 2,673 Japanese titles (774 ejournals, 29.0%)
  - 374 English titles (298 ejournals, 79.7%)

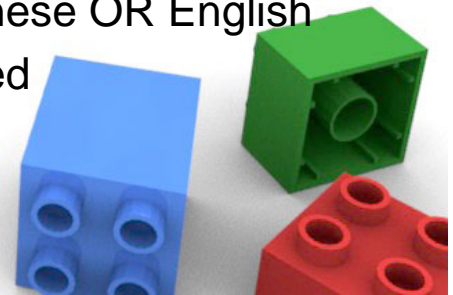
18

JATS-Con 2012



## Japanese STM Publishing (2)

- English-language journals are the same as Western ones in styles
- Japanese-language journals have:
  - Article titles, author names, affiliations – always in Japanese and English (Romanized)
  - Abstracts and keywords – in two languages mostly
  - Captions – either in Japanese or English
  - Cited references – either in Japanese OR English
  - Emphasis and Warichu are not used



19

JATS-Con 2012

### Example

#### 日本発行の科学技術分野の電子ジャーナル数

2005年から2008年への変遷

Electronic journal titles in science, technology and medicine published in Japan  
Changes from 2005 to 2008

時実 象一<sup>1</sup>

TOKIZANE Soichi<sup>1</sup>

<sup>1</sup> 愛知大学文学部 (〒441-8522 愛知県豊橋市町畑町1-1) Tel: 0532-47-4467 Fax: 0532-47-4197  
tokizane@u-aichi.ac.jp

<sup>1</sup> Faculty of Letters, Aichi University (1-1 Machihata-cho Toyohashi-shi, Aichi 441-8522)

原稿受理 (2010-12-27)

情報管理 54(1), 013-020, doi: 10.1241/johokanri.54.13 (<http://dx.doi.org/10.1241/johokanri.54.13>)

20

JATS-Con 2012



# Example

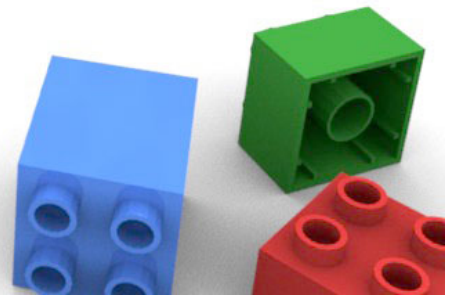
- Citations either in Japanese or English
  - Sometimes both for a single citation

7) 時実象一. 日本発行の人文社会系学術雑誌・紀要. 情報知識学会誌. 2008, vol. 18, no. 2, p. 204-208.

8) Tokizane, Soichi. Electronic Journal Publishing in Japan. World Conference on Agricultural Information and IT IAALD AFITA WCCA 2008. 2008-08-24/27, Tsukuba, <http://afita.ac.affrc.go.jp/wcca2008/index.htm>, (accessed 2010-12-27).

21

JATS-Con 2012

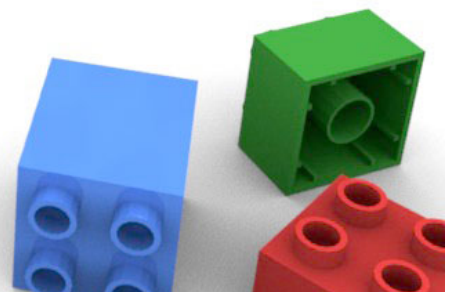


## Characteristics of Humanity/Social Science Publishing

- Vertical writing
  - Always for humanities
  - Often for social sciences
- Use of Empahsis and Warichu
- English-language article metadata are not common

22

JATS-Con 2012



# 中世の漢語「高名」「功名」について

李 妙 熙

キーワード 高名 功名 中世の漢語

要旨 漢語が日本語の中に定着する際、漢語とは多少異なる意味に転じる場合があるが、その傾向を見るために、中世の和漢語辞書などによく用いられている「高名」について、「功名」と関連して考察してみた。

その結果、「高名」のもとの意味は「有名」であるが、鎌倉初期に、「手柄(をたてる)」という意が生じ、中世全期を通じて用いられ、この意の「高名」が後の室町期に「功名」と混同されるようになり、江戸時代にも続くのであった。しかし、明治以降は、「高名」はもとの意である「有名」の意に用いられており、「手柄」の方は「功名」が用いられていることが観察できた。

## 一、はじめに

中世の漢語の特色の一つとしては、漢語の増大が注目

(東北大学大学院日本語学専攻 第1号 一九九一年)

これらは「手柄」の意としてとれる例であるが、「高名」の用例は中世にも少なくない。

また、「功名」の例も、室町後半に続いて見られる。

② かちはらとじゆんあみたふか二度のかけそれはかうみやうこれはみやうかう(きのふはけふの物語・下巻・45頁)

③ 俗も義臣すくつて此城にこもり、功名一時の義となる。(おくのほそ道・平泉・84頁)

右の例は、「手柄をたて、名をあげる」の意であって、江戸時代にも、「高名」と「功名」はほとんど同じ意として用いられている。『きのふはけふの物語』の例は、『平

される。中世になると、和漢語辞書の成立を見、漢語が増大し、多くの漢語が日本語の中に浸透し定着して行くのである。この定着の際、中には漢語の中で用いられた意味とは多少異なる意味に転じる場合も起こる。

そこで、本稿では、このような傾向の一端をうかがうために、和漢語辞書の異型といわれている『平家物語』『太平記』の中で用例がや多く見られる「高名」という語について検討していきたいと思う。

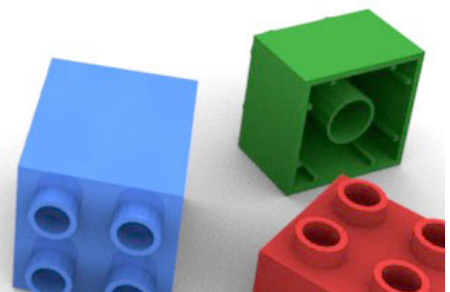
たとえば、①「いかに佐々木殿、高名せうどて不覚し給ふな。水の底には大綱あるらん」といひければ、『平家物語・下・巻九・四七頁]

②赤坂ノ大付金澤イコノ守太佛ノ奥州ニ向ツて宣ケル

ハ前日赤坂ノ城ヲメヲ奪シ候ツ事全ク土卒ノ高名ニアラス(太平記・巻七・47頁上)

## Ejournal Publishing in Japan

- Mostly in PDF
- Difficult to add value to ejournals
- XML is needed
  - not only for English-language journals
- Multi-lingual XML DTD is needed
  - to meet the Japanese-language journal needs to describe meta data in two languages



## Limitations of NLM DTD (3.0)

- Does not allow @xml:lang for authors and affiliations
- Bypassing example using @name-style

```
<contrib-group>
<contrib contrib-type="author">
<name name-style="western"><surname>Nihon</surname>
<given-names>Taro</given-names>
</name>
<name name-style="eastern"><surname>日本</surname>
<given-names>太郎</given-names>
</name>
</contrib>
```

25

JATS-Con 2012

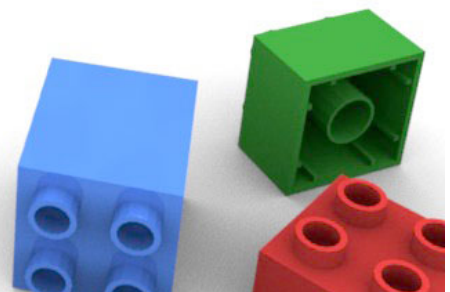


## Limitations of NLM DTD (3.0)

- Many elements not repeatable
  - <kwd-group>, <publisher-name>, etc.
  - unable to describe in multiple languages
- Some repeatable, but lack envelopes
  - <name>, <aff>
  - cannot indicate that descriptions in two languages belong to a single identity

26

JATS-Con 2012

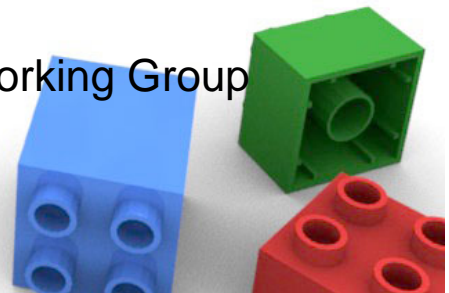


# Scholarly Publishing Japan (SPJ) Working Group

- Established by stakeholders in early 2010
  - Publishers
  - Typesetting companies
  - JST (observer)
  - NII (observer)
- Purpose
  - To enhance NLM DTD (JATS) so that it would fully support multi-language contents
  - Worked with NLM DTD (JATS) Working Group

27

JATS-Con 2012

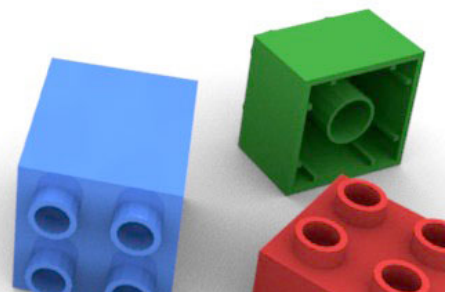


## SPJ Proposals (1)

- NLM DTD should be as much as structural
- Support both multiple languages as well as multiple scripts using IETF RFC 5646, for example `xml:lang="ja-Kana"`
- Stop using the default language as English for all elements
- Allow describing `@xml:lang` for `<name>`
- Introduce `<subbody>` to allow multiple language body texts

28

JATS-Con 2012

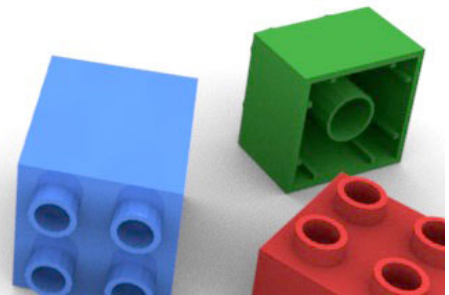


## SPJ Proposals (2)

- Devise ways to describe cited references in multiple languages, for example, <compound-element-citation>
- Allow describing <journal-meta> data such as <journal-title>, <journal-subtitle>, and <abbrev-journal-title> in multiple languages
- Allow specifying <xref> for individual <name> elements so that the institution name in the same script may be linked directly

29

JATS-Con 2012

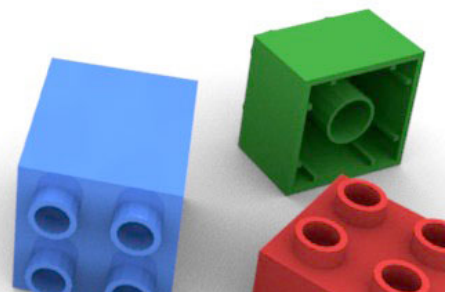


## NLM DTD Version 3.1 Draft

- 2010/9
  - Multi-lingual support
- 2011/3
  - Became NISO JATS 0.4

30

JATS-Con 2012



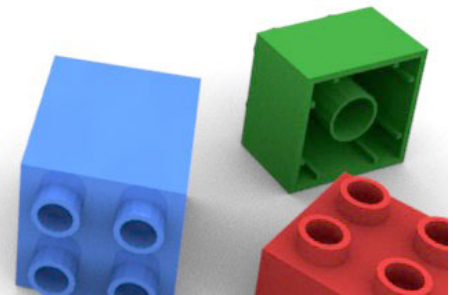


## Multi-lingual Support of JATS 0.4

- Deborah Aleyne Lapeyre and B. Tommie Usdin. Introduction to Multi-language Documents in JATS. Journal Article Tag Suite Conference (JATS-Con) Proceedings 2011. Sep. 26, 2011.

31

JATS-Con 2012

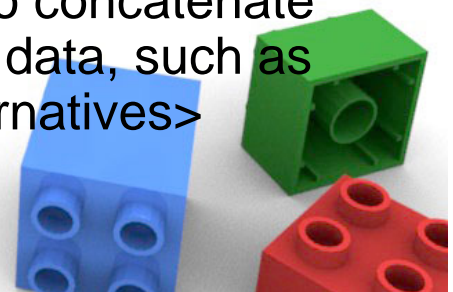


## Multi-lingual Support of JATS 0.4

- `xml:lang` is now usable for almost all the elements
- `xml:lang` is inherited down the XML document tree
- Both language and script may be recorded
- Most elements may be repeatable to describe multiple language expressions
- Wrapping tags are introduced to concatenate repeating elements for a single data, such as `<name-alternative>` or `<aff-alternatives>`

32

JATS-Con 2012



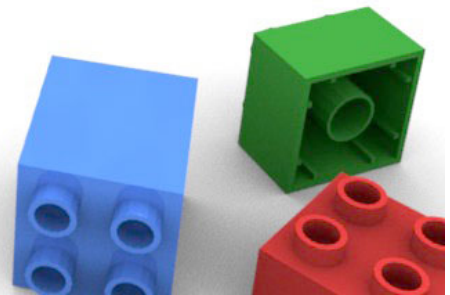


## @xml:lang with Scripts

- |                  |                               |
|------------------|-------------------------------|
| – 野田 佳彦          | xml:lang="ja-Jpan" (Kanji)    |
| – のだ よしひこ        | xml:lang="ja-Hira" (hirakana) |
| – ノダ ヨシヒコ        | xml:lang="ja-Kana" (katakana) |
| – Yoshihiko Noda | xml:lang="en" (Romaji)        |

33

JATS-Con 2012



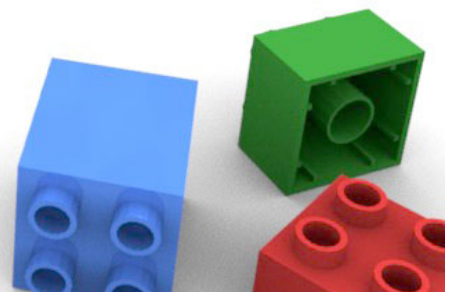
## Most Elements Repeatable

```
<kwd-group xml:lang="en">  
  <kwd>heated air</kwd>  
</kwd-group>  
<kwd-group xml:lang="ja">  
  <kwd>加温空気</kwd>  
</kwd-group>
```

- Except article titles
  - Use <trans-title>

34

JATS-Con 2012



## Wrapper

- <name-alternatives>

```
<name-alternatives>  
  <name name-style="eastern" xml:lang="ja-Jpan">  
    <surname>中西</surname>  
    <given-names>秀彦</given-names>  
  </name>  
  <name name-style="western" xml:lang="en">  
    <surname>Nakanishi</surname>  
    <given-names>Hidehiko</given-names>  
  </name>  
  <name name-style="eastern" xml:lang="ja-Kana">  
    <surname>ナカニシ</surname>  
    <given-names>ヒデヒコ</given-names>  
  </name>  
</name-alternatives>
```

35

## Wrapper

- <aff-alternatives>

```
<aff-alternatives>  
  <aff id="aff1">  
    <institution xml:lang="ja-Jpan">国立言語学博物館</institution>  
  </aff>  
  <aff id="aff-en">  
    <institution xml:lang="en">National Museum of  
    Linguistics</institution>  
  </aff>  
</aff-alternatives>
```

36

# Additional Proposals

- Coding Ruby

<rubigrp>多武峰<rubi>とうのみね</rubi></rubigrp>

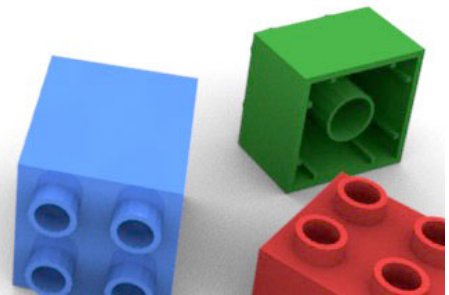
or

<ruby>多武峰<rt>とうのみね</rt></ruby>

(as in HTML5)

37

JATS-Con 2012



# Additional Proposals

- <ref-alternatives>

```
<ref-alternatives id="B30">
  <ref xml:lang="ja">
    <mixed-citation publication-type="journal" publication-format="print">
      <person-group person-group-type="author">
        <string-name name-style="eastern">
          <surname>柿崎</surname>
          <given-names>一郎 </given-names>
        </string-name>
      </person-group>. <year>2000</year>. 『<source>タイ 経済と鉄道 1835 年～1935
年</source>』 <publisher-name>日本経済評論社 </publisher-name>
    </mixed-citation>
  </ref>
  <ref xml:lang="en">
    <mixed-citation publication-type="journal" publication-format="print">
      <person-group person-group-type="author">
        <string-name name-style="western">
          <surname>Kakizaki</surname>, <given-names>Ichiro</given-names>
        </string-name>
      </person-group>. <year>2000</year>. <source>Thai Economy and Railway
1885-1935</source>. <publisher-loc>Tokyo</publisher-loc>:
<publisher-name>Nihon Keizai Hyoronsha</publisher-name>
    </mixed-citation> (in Japanese)
  </ref>
</ref-alternatives>
```

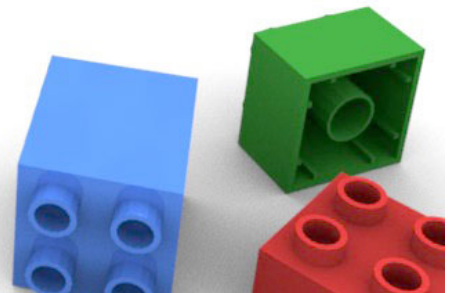


# New J-STAGE

- J-STAGE
  - E-journal platform operated by the Japan Science and Technology Agency (JST)
  - 2,387,426 articles/1,658 titles
- New J-STAGE launched in May, 2012
  - Supports JATS 0.4
- Several journals are loaded using JATS XML already

39

JATS-Con 2012



## 食道gastrointestinal stromal tumorに横隔膜上食道憩室を併発した1例

谷峰 直樹<sup>1)</sup>, 畑中 信良<sup>2)</sup>, 吉川 幸伸<sup>2)</sup>, 清水 洋祐<sup>2)</sup>, 遠藤 俊治<sup>2)</sup>, 西谷 暁子<sup>2)</sup>, 三隅 俊博<sup>2)</sup>, 中島 慎介<sup>2)</sup>, 上池 渉<sup>2)</sup>

1) 広島大学大学院医歯薬学総合研究科創性医学専攻先進医療開発科学講座外科学 2) 国立病院機構構内医療センター外科

J-STAGE公開日 2012/08/20

 本文PDF [2039K]

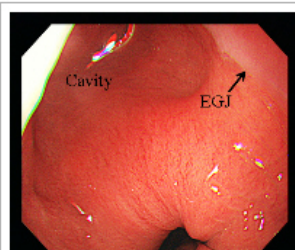
### Index

- ▼ Abstract
- ▼ はじめに
- ▼ 症例
- ▼ 考察
- ▼ 文献

### Abstract

症例は36歳の男性で、健診で腫瘍発生臓器の特定は困難で、に食道左側に突出する腫瘍を同様の検索にて中リスク群gastrointestinal stromal tumor (GIST) の存在が憩室形成に関与切除方法も個々の症例においてあった。我々は食道憩室を併発

上部消化管内視鏡検査所見: 腫瘍性病変は同定されず、胃食道接合部直上左側への内腔の突出を認めた。突出した内腔の観察は十分できなかったが、胃食道接合部胃粘膜の突出腔内への引き込み所見を認めたため、食道裂孔ヘルニアと診断した (Fig. 3)。



[View Details](#)

Fig. 3

Upper gastrointestinal endoscopy showed no submucosal tumor. It only showed a cavity overhanging the left side of the esophagogastric junction (EGJ). This part of the gastric mucosa was drawn into protuberant cavity. This was diagnosed as hiatal hernia.

### はじめに

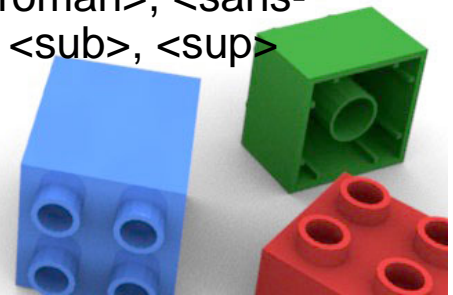
Gastrointestinal stromal tumor (GIST) は小腸およびその間膜を原発とする構造的・特異的な腫瘍である<sup>1)</sup>。また、術前に病理診断および我々は横隔膜上食道憩室を併発して報告する。

# J-STAGE XML Guideline (1)

- Characters
  - Characters are in UTF-8.
  - Entity references such as, &, <, >, &apos;, &quot;, and characters of ISO8879(SGML), MathML characters, and JATS specific characters such as &gcaron;, &Hmacr;, &euro;, and &franc;. may be used.
- Font attributes
  - <bold>, <italic>, <monospace>, <roman>, <sans-serif>, <sc>, <overline>, <strike>, <sub>, <sup> and <underline> may be used.

41

JATS-Con 2012

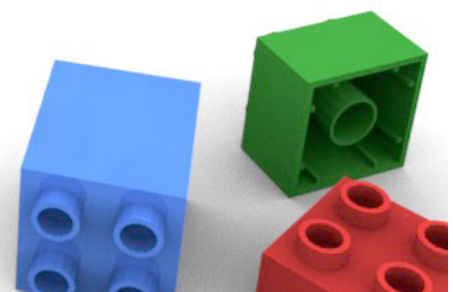


# J-STAGE XML Guideline (2)

- XML declaration and DOCTYPE
  - version="1.0" encoding="UTF-8"
  - <!DOCTYPE article PUBLIC "-//NLM//DTD JATS (Z39.96) Journal Publishing DTD v0.4 20110131//EN" "http://www.jstage.jst.go.jp/dtds/JATS-journalpublishing0.dtd">

42

JATS-Con 2012

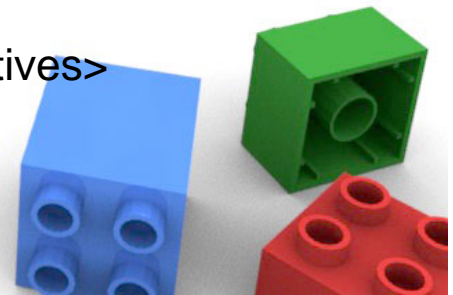


## J-STAGE XML Guideline (3)

- Journal meta
  - <journal-id> and <issn> must exist.
- Article meta
  - “publisher-id” should be JOI (JST Object Identifier, not publisher’s own id)
  - <contrib-group> must exist
  - <name-alternatives> and <aff-alternatives> required
  - <xref> id has to be in <aff-alternatives>

43

JATS-Con 2012

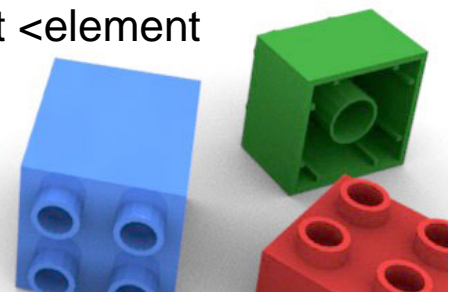


## J-STAGE XML Guideline (4)

- Mathematical formulas
  - Text, graphic, MathML and Tex/LaTeX are supported.
- Figures and Tables
  - Should appear where mentioned
  - OASIS/CALS not supported for a while
- References
  - <mixed-citation> is supported, not <element citation>.

44

JATS-Con 2012



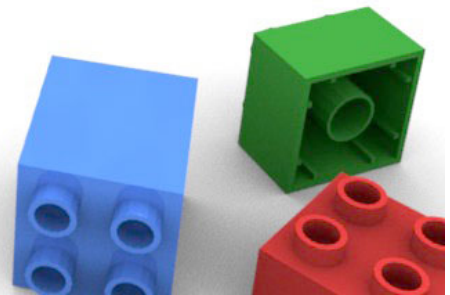


# Production Workflow (1)

- Typesetter A
  - MS Word 2007 to JATS XML (own program)
  - XML to XSL-FO
  - XSL-FO to PDF
  - XML Editor to make corrections

45

JATS-Con 2012

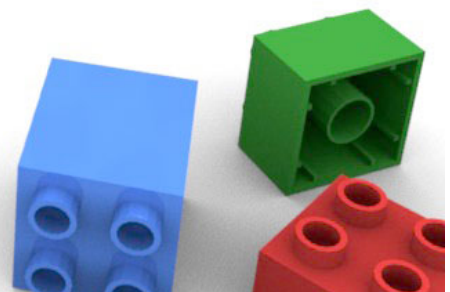


# Production Workflow (2)

- Typesetter B
  - TeX to JATS XML (own program)
  - MS Word to XML using eXtyle
    - Japanese-language XMLs need much human editing
  - XML to 3B2
  - 3B2 to JATS XML and PDF
  - Corrections on 3B2

46

JATS-Con 2012

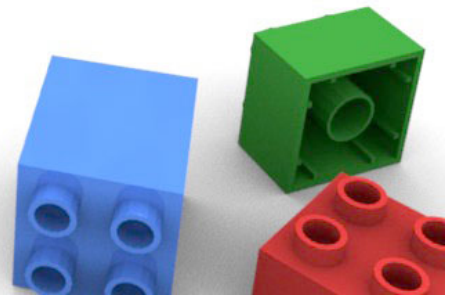


## Production Workflow (3)

- Typesetter C
  - Pastes texts to FrameMaker
  - FrameMaker-generated texts processed by eXtyle, then to JATS XML using XSLT
  - FrameMaker will be replaced by Typefi/InDesign soon

47

JATS-Con 2012

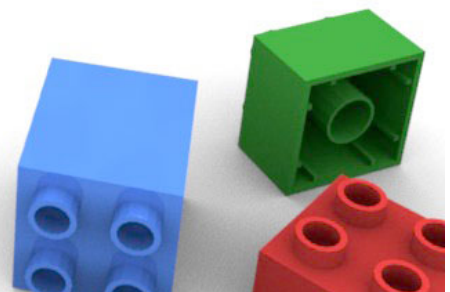


## Production Workflow (4)

- Typesetter D
  - MS Word to XML using eXtyle
    - Without cleanup for Japanese texts
  - XML to InDesign
  - InDesign XML to JATS XML using stylesheet
  - Corrections on InDesign

48

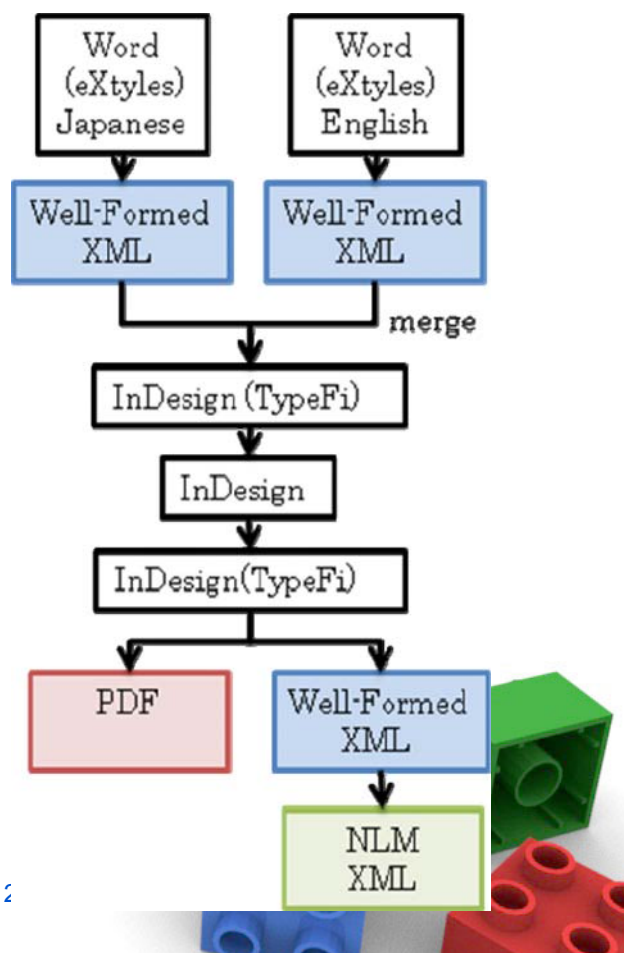
JATS-Con 2012





## InDesign/Typefi

- J. Comput. Chem. Jpn. 2011, 10(4), 141-146.



49

JATS-Con 2

## Problems and Issues with J-STAGE Guideline

- <name-alternatives> and <aff-alternatives> are must
- <element-citation> not supported
- "Publisher-id" should be a publisher's own id, rather than JOI.
- JST is working on them now.

50

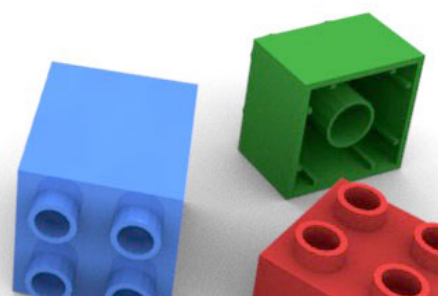
JATS-Con 2012

# Scholarly XML Publishing Association

- Extension of SPJ
- Established on June 28, 2012
- Soichi Tokizane named as President
- Symposium on September 19, 2012
- Future contact point of JATS Working group in Japan

51

JATS-Con 2012

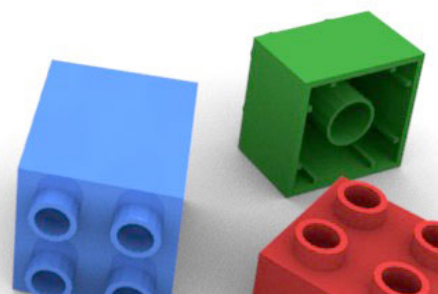


## What are next?

- J-STAGE has to be enhanced to fully take advantage of XML
- J-TAGE XML Guideline should be fixed
- Processing humanity/social science articles is very challenging
- Encourage scholarly book publishers to use JATS XML

52

JATS-Con 2012

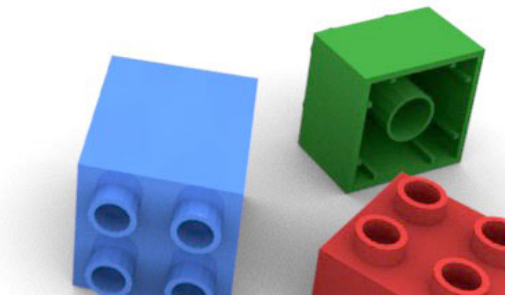


# Acknowledgement

- Soichi Kubota, JST
- Typesetters who gave me the information
- All SPJ members

53

JATS-Con 2012



# Questions?

- Soichi Tokizane  
– tokizane@aichi-u.ac.jp

54

JATS-Con 2012

