Everybody knows what metadata is!

It is data about data!

MTHFR methylenetetrahydrofolate reductase [ Homo sapiens (human) ]
Official Symbol: MTHFR provided by HGNC
Official Full Name: methylenetetrahydrofolate reductase provided by HGNC
Primary source: HGNC:HGNC:7436
See related: Ensembl:ENSG00000177000; MIM:607093; Vega:OTTHUMG00000002277
Gene type: protein coding
RefSeq status: REVIEWED
Organism: Homo sapiens
Lineage: Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Summary: The protein encoded by this gene is deficient [provided by RefSeq, Oct 2009]
Expression: Ubiquitous expression in lung (RPKM 7.5), thyroid (RPKM 7.2) and 25 other tissues
See more
Orthologs: mouse all
ipmc-dev11$ identify -verbose buttercup.jpg
Image: buttercup.jpg
Format: JPEG (Joint Photographic Experts Group JFIF format)
Class: DirectClass
Geometry: 487x517+0+0
Resolution: 72x72
Print size: 6.76389x7.18056
Units: PixelsPerInch
Type: TrueColor
Endianness: Undefined
Colorspace: sRGB
Depth: 8-bit
Channel depth:
red: 8-bit
green: 8-bit
blue: 8-bit
Channel statistics:

**Red:**
- min: 0 (0)
- max: 255 (1)
- mean: 133.867 (0.524968)
- standard deviation: 72.416 (0.283984)
- kurtosis: -1.08949
- skewness: 0.217215

**Green:**
- min: 0 (0)
- max: 255 (1)
- mean: 92.3975 (0.362343)
- standard deviation: 65.9452 (0.258609)
- kurtosis: -0.431805
- skewness: 0.732417

**Blue:**
- max: 255 (1)
- mean: 71.8805 (0.281884)
- standard deviation: 58.3456 (0.228806)
- kurtosis: 0.332941
- skewness: 1.07552

Image statistics:

**Overall:**
- min: 0 (0)
- max: 255 (1)
- mean: 99.3816 (0.389732)
- standard deviation: 65.8206 (0.25812)
- kurtosis: 0.158924
- skewness: 0.814258

**Rendering intent:** Perceptual

**Gamma:** 0.454545

**Chromaticity:**
- red primary: (0.64,0.33)
- green primary: (0.3,0.6)
- blue primary: (0.15,0.06)
- white point: (0.3127,0.329)

**Interlace:** None

**Background color:** white

**Border color:** rgb(223,223,223)

**Matte color:** gray74

**Transparent color:** black

**Compose:** Over

**Page geometry:** 487x517+0+0

**Dispose:** Undefined

**Iterations:** 0

**Compression:** JPEG

**Quality:** 95

**Orientation:** TopLeft
Everybody knows what metadata is!

It is data about data!
Everybody knows what metadata is!

It is any information about data!

Everybody knows what metadata is!

But it can also be about things.

Metadata

Name: Buttercup
Species: Dog (Doggis doggis)
Breed: Mutt
Age: 8 ½ years
Color: Brown (with some gray)
Who doesn't know what this is?

Metadata Inside!

But what exactly?
But what exactly?
The Work?

Work - “a distinct intellectual or artistic creation”

Detour to FRBR
Functional Requirements for Bibliographic Records by the IFLA.
Google “FRBR”
Section 3.2 “The Entities”

Work - “a distinct intellectual or artistic creation”

Expression - “the intellectual or artistic realization of a work in the form of alphabetic, musical, or choreographic notation, sound, image, object, movement, etc., or any combination of such form”

Detour to FRBR
Functional Requirements for Bibliographic Records by the IFLA.
Google “FRBR”
Section 3.2 “The Entities”
Work - “a distinct intellectual or artistic creation”

Expression - “the intellectual or artistic realization of a work in the form of alphanumeric, musical, or choreographic notation, sound, image, object, movement, etc., or any combination of such form”

Manifestation - “the physical embodiment of an expression of a work.”

Item - “a single exemplar of a manifestation”

Detour to FRBR

Functional Requirements for Bibliographic Records by the IFLA.

Google “FRBR”

Section 3.2 “The Entities”

But what exactly?
The Work?
The Expression?
The Manifestation?
The Item?
It is any information about anything!

And if you are “talking metadata” with someone, you both need to agree on the subject and the properties that you are working with (and most likely what the properties will be used for).
JATS is a NISO standard that defines XML elements and attributes and models for describing Journal Articles.

Another Detour

JATS XML
Like any good story

An article represented in JATS XML has 3 parts.
Like any good story

An article represented in JATS XML has 3 parts.

Identifying and Locating metadata are kept in the <front>.

Expanding <front>
Bee Threat Elicits Alarm Call in African Elephants

African elephants respond to the presence of bees with an alarm call. This behavior may serve as a warning to other elephants about potential danger, as bees can deliver painful stings. Observations indicate that elephants exhibit an acute response to bees, showing signs of distress and vocalizing when stung. This suggests a level of awareness and social integration in which elephants are sensitive to non-human threats, enhancing their ability to navigate and protect their habitat.
Conceived and designed the experiments: LEK JS IDH AS FV. Performed the experiments: LEK JS. Analyzed the data: LEK JS. Wrote the paper: LEK JS IDH AS FV.
In JATS, an article must have either a First Page number (<fpage>) or a unique-in-volume identifying string (<elocation-id>) - sometimes thought of as an electronic page number.

The <elocation-id> stands in for the page number in an article citation.
The article citation is a name (or alternate name) for the article that identifies the article


This is MetaData in action!
Problematic Metadata Elements?

Permissions/Licensing Information
Publication Dates
Funding Information
Conflict of Interest Statements
Problematic Metadata Elements?
They are all “easy” to tag!

<permissions>
<copyright-statement>© 2014 Surname et al.</copyright-statement>
<copyright-year>2014</copyright-year>
<copyright-holder>Surname et al.</copyright-holder>
<license>
<ali:free_to_read/>
<license-p>This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, reproduction and adaptation in any medium and for any purpose provided that it is properly attributed. For attribution, the original author(s), title, publication source (PeerJ) and either DOI or URL of the article must be cited.</license-p>
</license>
</permissions>

Problematic Metadata Elements?
They are all “easy” to tag!

<pub-date publication-format="print" date-type="pub" iso-8601-date="1999-01-29">
<day>29</day>
<month>01</month>
<year>1999</year>
</pub-date>
<pub-date publication-format="electronic" date-type="original-publication" iso-8601-date="2018-01-29">
<day>29</day>
<month>01</month>
<year>2018</year>
</pub-date>
<pub-date publication-format="electronic" date-type="update" iso-8601-date="2018-05-07">
<day>07</day>
<month>05</month>
<year>2018</year>
</pub-date>

Problematic Metadata Elements?
They are all “easy” to tag!

<funding-group specific-use="Crossref">
<award-group>
<funding-source id="gs1" country="US">
<institution-wrap>
<institution>National Institutes of Health</institution>
</institution-wrap>
<institution-id institution-id-type="doi" vocab="open-funder-registry" vocab-identifier="10.13039/open_funder_registry">10.13039/100000002</institution-id>
</funding-source>
<award-id>GM18458</award-id>
</award-group>
<award-group>
<funding-source id="gs2" country="US">
<institution-wrap>
<institution>National Science Foundation</institution>
</institution-wrap>
<institution-id institution-id-type="doi" vocab="open-funder-registry" vocab-identifier="10.13039/open_funder_registry">10.13039/100000001</institution-id>
</funding-source>
<award-id>DMS-0204674</award-id>
<award-id>DMS-0244638</award-id>
</award-group>
</funding-group>
Problematic Metadata Elements?

They are all “easy” to tag!

Tagging these metadata is
not an issue

Using your metadata in your publishing system - where you control the creation/tagging and the use of the content - will not be a problem.

Challenges arise when someone else tries to use your metadata.

Tagging these metadata is
not an issue

These items are on the “problematic” list because they contain information that others want to use from/know about your article:

1. What can/can’t they do with the article?
2. When was it published?
3. Who paid for it?
4. More of who paid for it?
Now some slides from a 2015 presentation at JATS-Con, “Improving the reusability of JATS”

https://www.ncbi.nlm.nih.gov/books/NBK279901/
If you think no one will reuse your content

Think again!

Aggregators, archives, libraries, indexing services

But the biggest reuser of your content will most likely be you!

Metadata is any information about anything.

Understand when you are talking "MetaData" that you need to define:
1. WHAT information about WHAT
2. To be used to do WHAT

Defining and tagging metadata items consistently now ...
will make life easier for future editors, text miners, and researchers.