



Universität Leipzig

Fakultät für Biowissenschaften, Pharmazie und Psychologie

Meta data in research

Marc Schönwiesner

Schleyer-Seminar

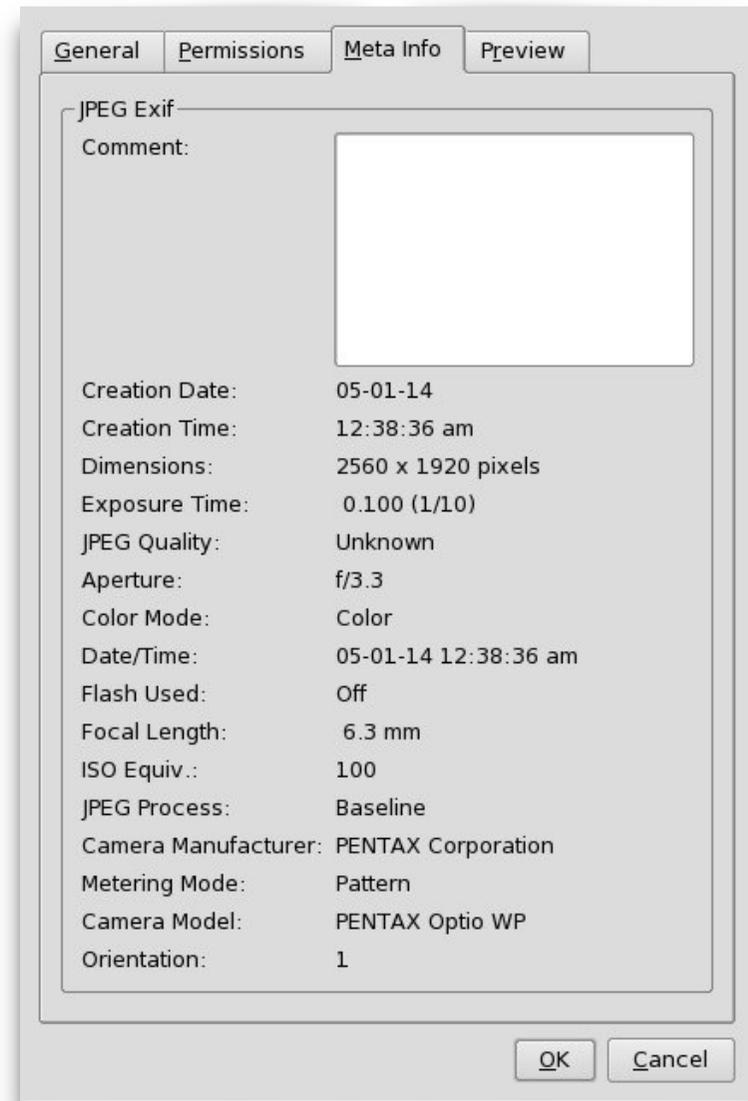
Meta data

It's everywhere...

Descriptive (title, abstract, author, keywords)

Structural (data containers, page order)

Administrative (date created, changed, access)



Format: Abstract

Send to

Hist Psychol. 2017 Aug;20(3):342-345. doi: 10.1037/hop0000068.

The digitization of the Wundt estate at Leipzig University.

Meyer T¹, Mädebach A¹, Schröger E¹.

Author information

Abstract

Wilhelm M. Wundt (1832-1920) was one of the most important German scholars of the 19th and early 20th centuries and famously founded the first institute for experimental psychology in Leipzig in 1879. Wundt's institute established a teaching and research facility that attracted a large number of students from all over the world and contributed greatly to the development of modern psychology. Until now, the relatively poor indexing and documentation as well as the difficulty in accessing the Wundt estate has prevented a widespread and comprehensive investigation and consideration of these documents. The digitization project described in this article has rectified these problems and will hopefully provide a valuable source for students and researchers interested in Wundt's work. (PsycINFO Database Record.

PMID: 28782973 DOI: 10.1037/hop0000068



LinkOut - more resources



Full text links



Save items

Add to Favorites

Similar articles

The emergence and development of Bekhterev's psychoreflexology in rela [J Hist Behav Sci. 2014]

Wundt's laboratory at Leipzig in 1891. [Hist Psychol. 1999]

[Reaction time tests in Leipzig, Paris and Würzburg. The Franco-Gern [Medizinhist J. 2004]

Review Merging with the path not taken: Wilhelm Wundt's work as [Conscious Cogn. 2018]

Review Chapter 33: the history of movement disorders. [Handb Clin Neurol. 2010]

See reviews...

See all...



Cataloging and Metadata Management

Home > About the NLM > Cataloging and Metadata Management

NLM Metadata Schema

(Approved February 2002)

(Revised July 29, 2004)

The NLM metadata schema is based on the metadata terms maintained by the Dublin Core Metadata Initiative, including elements, element refinements, encoding schemes, and vocabulary terms (cf. <http://dublincore.org/documents/dcmi-terms/>). It also incorporates some of the best practice recommendations of the DC-Library Application Profile (DC-Lib) (cf. <http://dublincore.org/documents/2002/04/16/library-application-profile/>). It is designed for use with electronic resources published by the Library and incorporates additional elements and qualifiers identified as requirements by NLM, most notably the addition of Permanence ratings for each resource. ¹

In the list which follows, the elements in the NLM Metadata schema are outlined using the following conventions:

The key used for "Identifier" is:

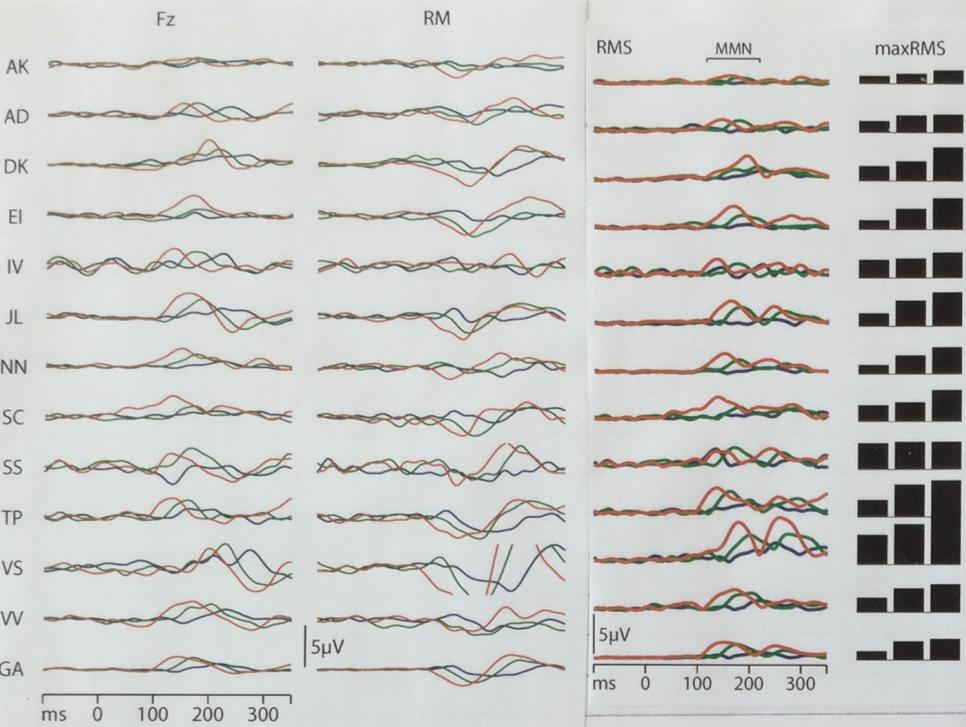
- DC = Approved Dublin Core elements and qualifiers**
- NLMDC = Approved Dublin Core elements with NLM-defined qualifiers**
- NLM = NLM-defined elements**

The key for the element requirements is:

- R = Required**
- RA = Required if applicable**
- O = Optional**

Element: Title

Name: Title
Identifier: DC.Title



EEG data ↑ Fz (Cz), right mastoid, and the maximum of RMS is compared to individual effects.

Table 1
Some central characteristics of auditory MMN

Independence of attention
MMN is best observed when the subject's attention is directed away from the auditory stimuli, as the overlap of other negative components at the same latency range, (e.g. N2b) is avoided
It reflects preattentive change detection because it is elicited even when participants perform a task that is not related to the auditory stimuli
It is not completely attention-independent in all conditions, but it is not abolished by the withdrawal of attention

Endogeneity
The latency of MMN is inversely related to, and its amplitude positively related to the magnitude of the difference between the standard and the deviant stimulus
There is a relationship between MMN latency and reaction time, since both diminish when the physical difference between standard and deviant stimulus is increased

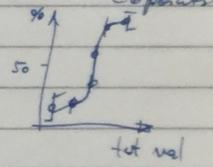
Sensory memory
The MMN provides an index of experience-dependent memory traces in the human brain
It is a response to the relation between the present stimulus and the previous stimulus. It is elicited by infrequent violations of acoustic regularities
It has been recently suggested to reflect long-term memory traces for language sounds such as phonemes and syllables

Sensory discrimination
MMN provides an index of the perceptual discrimination accuracy for different simple and complex sound features
It is elicited by any discriminable change of a repetitive sound or regularity in sound sequences
It may be elicited by stimulus differences that approximate the behavioral discrimination threshold

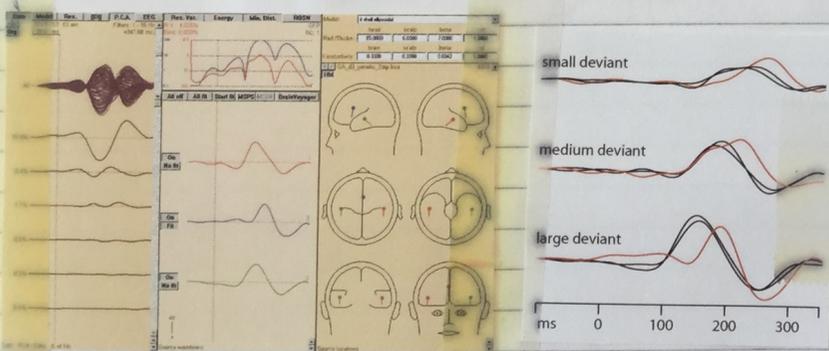
Cerebral sources
The MMN has a bilateral auditory-cortex generator (auditory supratemporal cortex) and a right frontal cortex generator, among others
It has been observed that for deviances in intensity, frequency and duration, the MMN is larger over the right hemisphere regardless of the stimulated ear

Attentional reorienting
It has been proposed that the cerebral discrimination process generating the MMN may play an important role in involuntary orienting, or attention switching to a change in the acoustic environment

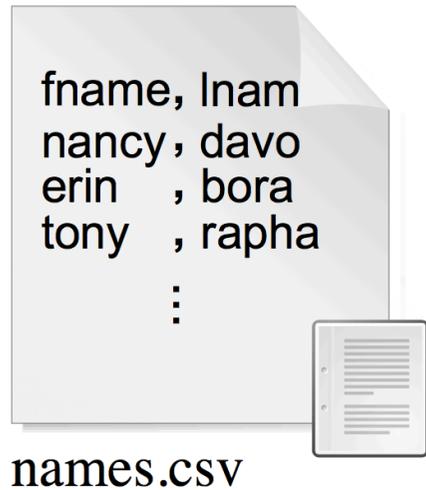
Smith et al. 05 JdA: "processing of size info"
single vowel paradigm: 2I-2AFC → which is smaller (greater)?
speed-like — : play a train of stimuli in eds intervals
60 trials / psychofet. (6 points)



only constant feature in interval: size,
only constant data between interv.: —
→ 6pt psychofet, 10 trials per point
fit cumulative Gaussians (Foster & Bischof 97)
calculate jnd from fitted curve:
can be defined as diff. in test value for 50% and 76% correct
(d'=1 in this 2AFC task) rel. to perceived test value (50%)



Plain Text and CSV



PAReverb ipt

NRuns	4			
_refRoomSizes	12	128	256	512
_firstSeps	20	64	128	256
_iStepsUp	4	16	24	48
_iStepsDown	4	16	24	48
_fStepsUp	2	4	8	16
_fStepsDown	2	4	8	16
NDown	3			
NUp	2			
paradigm	3d2u			
NPTrials	4			
NTurnPoints	6			
catname	christoscat			

XML and JSON

JSON Example

```
{ "employees": [
  { "firstName": "John", "lastName": "Doe" },
  { "firstName": "Anna", "lastName": "Smith" },
  { "firstName": "Peter", "lastName": "Jones" }
]}
```

XML Example

```
<employees>
  <employee>
    <firstName>John</firstName> <lastName>Doe</lastName>
  </employee>
  <employee>
    <firstName>Anna</firstName> <lastName>Smith</lastName>
  </employee>
  <employee>
    <firstName>Peter</firstName> <lastName>Jones</lastName>
  </employee>
</employees>
```



MEDLINE[®]/PubMed[®] XML Data Elements

This document reflects the [DTD](#) used for the 2018 MEDLINE/PubMed production year.

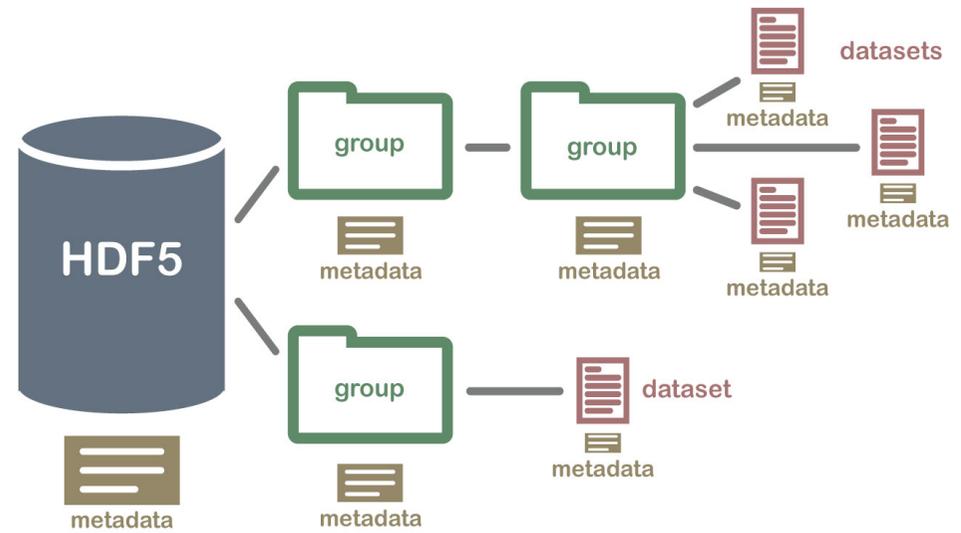
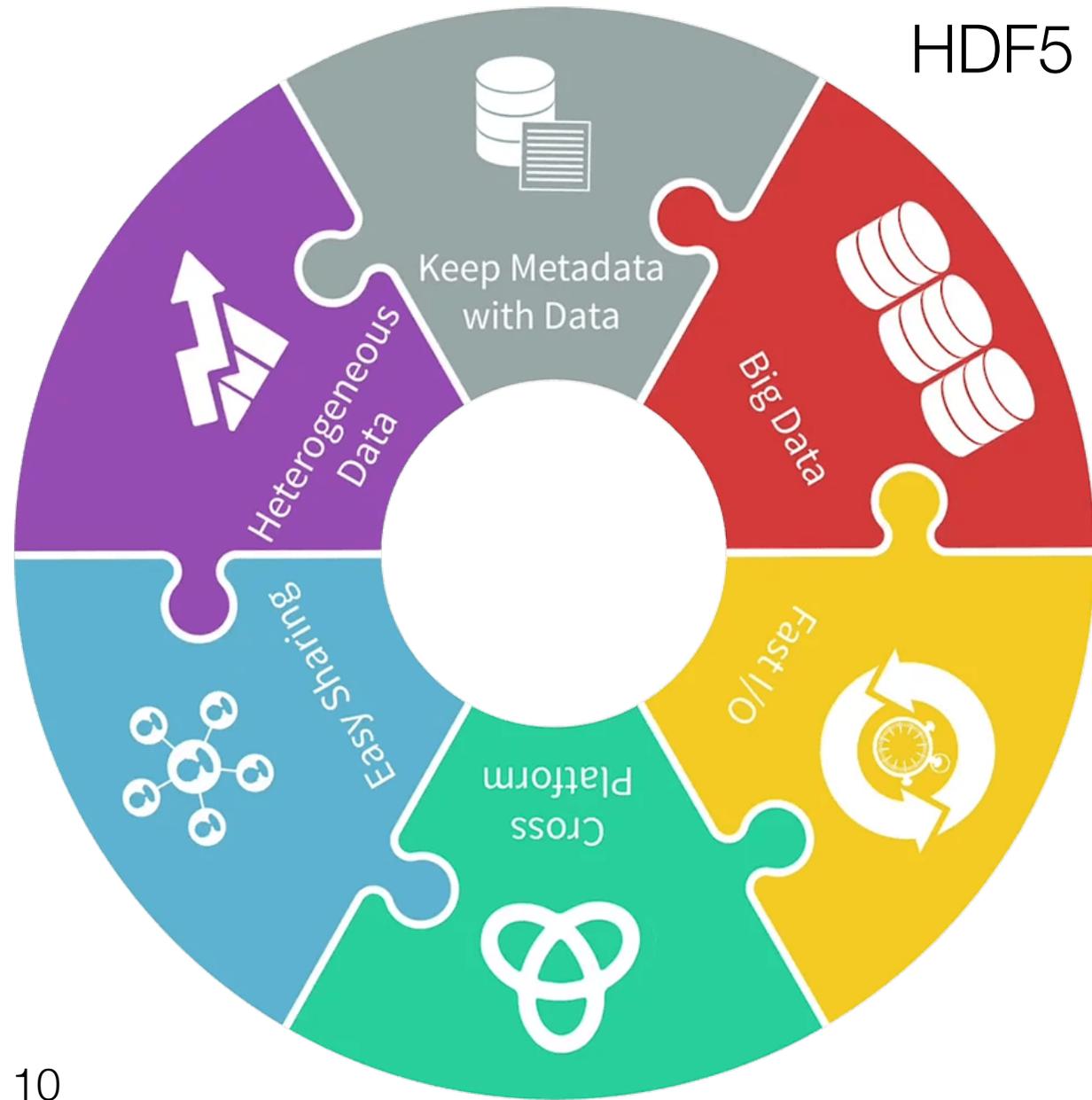
This document is prepared for those who lease MEDLINE/PubMed from NLM and other interested parties.

- [XML Element Descriptions and their Attributes](#)
 - [Alphabetical List of the Elements](#)
 - [MedlineCitationSet \(MedlineCitation, DeleteCitation\)](#)
- [Creation of Journal Source at NLM](#)

The issue of combining data and meta data



HDF5



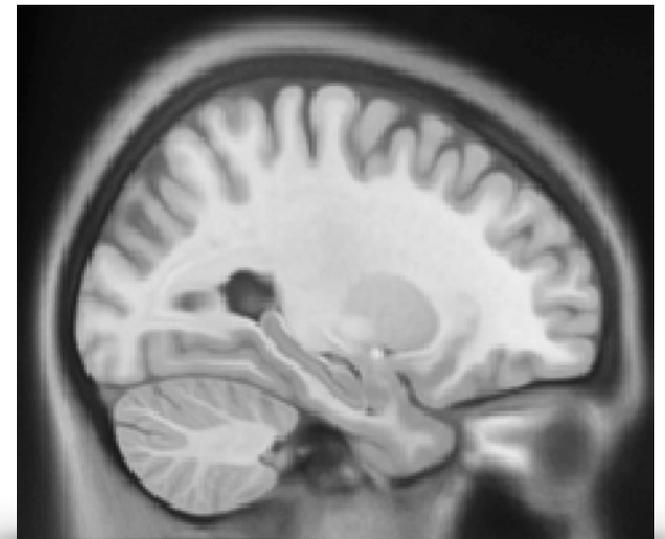
```
46 class AttributeTypes:
47     Conventions = 0 # Specifies the netCDF file as a set of AES-X212 SOFAConventions.
48     Version = 1 # Version of the AES-X212 specifications. The version is in the form x.y, where x is the version major and y the vers
49     SOFAConventions = 2 # Name of the AES-X212 convention.
50     SOFAConventionsVersion = 3 # Version of the AES-X212 convention. The version is in the form x.y, where x is the version major and
51     DataType = 4 # Specifies the data type
52     RoomType = 5 # Specifies the room type.
53     Title = 6 # A succinct description of what is in the file.
54     DateCreated = 7 # Date and time of the creation of the file in ISO 8601 format: 'yyyy-mm-dd HH:MM:SS'. This field is updated each
55     DateModified = 8 # Date and time of the last file modification in ISO 8601 format: 'yyyy-mm-dd HH:MM:SS'. This field is updated e
56     APIName = 9 # Name of the API that created/edited the file
57     APIVersion = 10 # Version of the API that created/edited the file. The version is in the form x.y, where x is the version major a
58     AuthorContact = 11 # Contact information (for example, email) of the author
59     Organization = 12 # Legal name of the organization of the author. Use author's name for private authors
60     License = 13 # Legal license under which the data are provided
61     ApplicationName = 14 # Name of the application that created/edited the file
62     ApplicationVersion = 15 # Version of the application that created/edited the file
63     Comment = 16 # Miscellaneous information about the data or methods used to produce the date/file
64     History = 17 # Audit trail for modifications to the original data
65     References = 18 # Published or web-based references that describe the data or methods used to produce the data
66     Origin = 19 # The method used for creating the original data. In case of model-generated data, origin should name the model and i
67     RoomShortName = 20 # Short Name of the room
68     RoomDescription = 21 # Informal verbal description of the room
69     RoomLocation = 22 # Location of the room
70     ListenerShortName = 23 # Short name of the listener
71     ListenerDescription = 24 # Description of the listener
72     SourceShortName = 25 # Short name of the source
73     SourceDescription = 26 # Description of the source
74     ReceiverShortName = 27 # Short name of the receiver
75     ReceiverDescription = 28 # Description of the receiver
76     EmitterShortName = 29 # Short name of the emitter
77     EmitterDescription = 30 # Description of the emitter
```


Standard MINC Variable Names^[edit]

- *image* - The *image* variable is of class *group*. It is the variable which actually contains image data in a MINC file, so it is defined using whatever type and dimensions are required to represent the image. By convention, MINC considers the first spatial dimension of the *image* variable to be the 'slice' dimension, and any other spatial dimensions are considered to be 'image' dimensions. The *image* variable is the only variable whose presence in a MINC file is mandatory.
- *image-min* - The *image-min* variable is of class *var-attribute*. If per-slice scaling of the image data is enabled, this variable must contain 64-bit floating-point data which provides the lowest value of the real range for each slice. The dimensionality of either the *image-min* or the *image-max* variable corresponds to the first one or two dimensions of the *image* variable. For example, in an fMRI dataset with a dimension ordering of *time*, *zspace*, *yspace*, *xspace*, the *image-min* variable may be a 2D array with dimensions *time* and *zspace*. It is acceptable for the *image-min* to have lower dimensionality. For example, the variable could be a 1D array array along the *time* dimension, or it may be a scalar to specify a global minimum real value.
- *image-max* - The *image-max* is the counterpart to the *image-min* variable, except that it contains the maximum value of the real data range for each data slice.
- *study* - This variable is of class *group*. It contains no useful data, but serves only to group those attributes which contain information about the study of which this image is a part. Since the variable contains no data, the type is irrelevant.
- *patient* - Like the *study* variable, this variable is of class *group* and contains no data. It serves to group the attributes which specify the identification and characteristics of the patient.
- *acquisition* - The *acquisition* variable is of class *group*. Like the *study* and *patient* variables, the *acquisition* variable never contains useful data, but serves only to group those attributes which contain

Example MINC format (H5)

```
1 CDF
2 zspace yspace xspace
3 ident+bouffard:oliver:2010.11.15.15.29.12:2167:1
4 history
5 Mon Nov 15 15:29:12 2010
6 >>> mincmath -add PACN10_TE10_b_nlin2StdMNI.mnc PACN10_TE11_b_nlin2StdMNI.mnc PACN10_TE12_b_nlin2StdMNI.mnc PACN10_TE_sum.mnc
7 zspacevarid MINC standard variable vartype dimension__ version MINC Version 1.0 comments Z increases from patient inferior to superiorspacing
8 yspacevarid MINC standard variable vartype dimension__ version MINC Version 1.0 comments Y increases from patient posterior to anteriorspacing
9 xspacevarid MINC standard variable vartype dimension__ version MINC Version 1.0 comments X increases from patient left to rightspacing
10 image-maxvarid MINC standard variable vartype var_attribute version MINC Version 1.0
11 imageparent root variable image-max--->image-max image-min--->image-mincompletetrue_signtype unsigned 1.0parentchildrenimage
12 ?@##### @!#####@$$c$@*a+#####@,a-#####@+#####@.a-#####@.a@
```



Anonymising

The screenshot shows the DicomBrowser application interface. The 'Edit' menu is open, with 'Clear' highlighted. The background shows a tree view of DICOM data and a table of attributes.

File Tree:

- Patient Sample ID
 - Study 1
 - MR Series 4
 - Image 1 : /Use
 - Image 2 : /Use
 - Image 3 : /Use
 - Image 4 : /Use
 - Image 5 : /Use
 - Image 6 : /Use
 - Image 7 : /Use
 - Image 8 : /Use
 - Image 9 : /Use
 - Image 10 : /Us
 - Image 11 : /Us
 - Image 12 : /Us
 - Image 13 : /Us
 - Image 14 : /Us
 - Image 15 : /Us

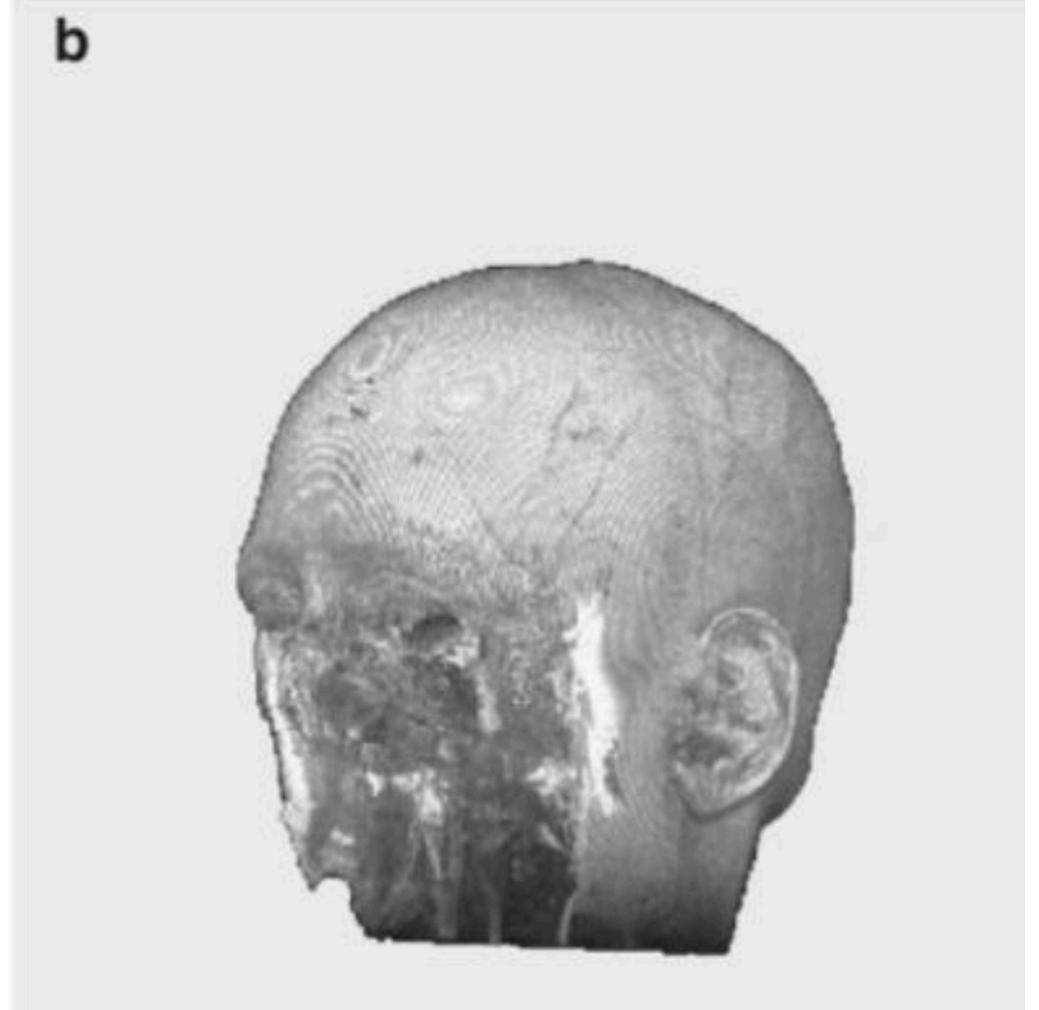
Menu Items:

- Undo (⌘Z)
- Redo (⌘Y)
- Keep
- Clear**
- Delete
- Add new attribute... (⌘I)
- Apply script...

Table of Attributes:

Tag	Attribute Name	Action	Value
(0008,1030)	Study Description	Keep	4 values: 093711.890000, 093711.906000, 0...
(0008,103E)	Series Description	Keep	.
(0008,1050)	Performing Physicia...	Keep	MR
(0008,1090)	Manufacturer's Mo...	Keep	SIEMENS
(0008,1140)	Referenced Image ...	Keep	Hospital
(0010,0010)	Patient's Name	Keep	StreetStreetNo, City, District, US, ZIP
(0010,0020)	Patient ID	Keep	an'...
(0010,0030)	Patient's Birth Date	Keep	MEDPC
(0010,0040)	Patient's Sex	Keep	head^DHead
(0010,1010)	Patient's Age	Keep	t1_mpr_1mm_p2_pos50
(0010,1030)	Patient's Weight	Keep	TrioTim
(0018,0015)	Body Part Examined	Keep	{sequence}

Anonymising



Gene [Advanced](#) [Help](#)

Full Report [Send to:](#)

INS insulin [*Homo sapiens* (human)]

Gene ID: 3630, updated on 2-Oct-2018

Summary ↑ ?

- Official Symbol** [INS](#) provided by [HGNC](#)
- Official Full Name** [insulin](#) provided by [HGNC](#)
- Primary source** [HGNC:HGNC:6081](#)
- See related** [Ensembl:ENSG00000254647](#) [MIM:176730](#); [Vega:OTTHUMG00000009558](#)
- Gene type** protein coding
- RefSeq status** REVIEWED
- Organism** [Homo sapiens](#)
- Lineage** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
- Also known as** [IDDM](#); [ILPR](#); [IRDN](#); [IDDM1](#); [IDDM2](#); [MODY10](#)
- Summary** After removal of the precursor signal peptide, proinsulin is post-translationally cleaved into three peptides: the B chain and A chain peptides, which are covalently linked via two disulfide bonds to form insulin, and C-peptide. Binding of insulin to the insulin receptor (INSR) stimulates glucose uptake. A multitude of mutant alleles with phenotypic effects have been identified. There is a read-through gene, INS-IGF2, which overlaps with this gene at the 5' region and with the IGF2 gene at the 3' region. Alternative splicing results in multiple transcript variants. [provided by RefSeq, Jun 2010]
- Expression** Restricted expression toward pancreas (RPKM 671.7) [See more](#)
- Orthologs** [mouse](#) [all](#)

Genomic context ↑ ?

Location: 11p15.5 [See INS in Genome Data Viewer](#)
Exon count: 3

Annotation release	Status	Assembly	Chr	Location
--------------------	--------	----------	-----	----------

Our own solution, based on HDF5

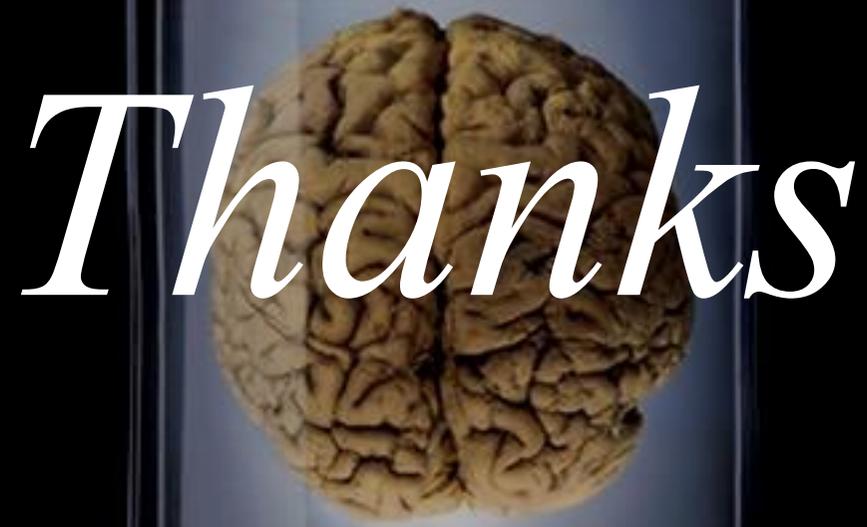
```
- experiment (tables.Group), title = experimentName_Nr # node for one experiment
  _v_attrs (tables.AttributesSet) # this is the metadata for the experiment
    - start_time (str)
    - end_time (str)
    - computer (str)
    - experimenter (str)
    - status (str)
    - Global_setting (str, all global parameters)
    - experiment_setting (str, parameters in model.setting)
    - device_setting (str, parameters in device.setting)
    ...

**trial_0000 (tables.Group), title = trial_0000**
  _v_attrs (tables.AttributesSet) # this is the metadata for the trial
    - trial related settings
    - start_time
    - end_time

| event_log (tables.Table)
| trial_log (tables.Table)
| data_00 (tables.EArray)
| data_01 (tables.EArray)
```

Takehome

- metadata is necessary to structure and document your research data
- you are already using it everyday
- use industry standard file formats, but a lab book will do in a pinch
- force the entry (otherwise no one will do it)
- but make most of it automatic (otherwise no one will do it)



Thanks