

NWB (.nwb)

Legend

comments

* new section or null/unknown value or delimiter between the attribute name and the data type

() optional

Preserved Key Words (case sensitive)

- *Nodes
- *DirectedEdges
- *UndirectedEdges
- id
- source
- target
- weight
- *int
- *string
- *float

NWB Format

```
# Comments
*Nodes (total numbers of nodes)
id*int      label*string      (nodeAttribute1*data_type)      (nodeAttribute2*data_type) ...
...
*DirectedEdges (total numbers of edges)
source*int   target*int      (edgeAttribute1*data_type)      (edgeAttribute2*data_type)
...
*UndirectedEdges (total numbers of edges)
source*int   target*int      (edgeAttribute1*data_type)      (edgeAttribute2*data_type)
...
```

Example

```
# This is a paper-author network.
# Author labels are author names.
# Paper labels are titles.
# Paper weights indicate # citations.
*Nodes 4
id*int  label*string      weight*int  node_type*string
1       "Joe Ann"           0          "author"
2       "John Smith"      0          "author"
3       "Bio Today"       8          "paper"
4       "Physics Tomorrow" 15         "paper"
*DirectedEdges 2
source*int  target*int  weight*float  edge_type*string
1           3           0.66         "wrote"
4           3           0.78         "paper-citation"
```

Detailed Specification

1. There are three types of section headers and they are case sensitive:
 - *Nodes
 - *DirectedEdges
 - *UndirectedEdges
2. There are three data types, string, integer, and floating point.
3. The *Nodes section header MUST NOT be omitted in the NWB file. The "total numbers of nodes" after *Nodes is optional, and should be separated only by arbitrary number of spaces and/or tabs.
4. The list of nodes under the section of *Nodes MUST NOT be omitted and MUST provide a complete list including the node id and node label. The data type of node ids must be integer and the value must be bigger than 0; in another word, the minimum node id is 1). The data type of node labels must be string.
5. One of *DirectedEdges section or *UndirectedEdges section is required. If desired, both may be used, but at least one must be. A graph with non-empty *DirectedEdges and *UndirectedEdges sections is a hybrid graph. "Total numbers of edges" is optional after *DirectedEdges or *UndirectEededges, and should be separated only by arbitrary number of spaces and/or tabs.
6. All edges must be listed, none are inferred.

7. All string values (not including the attribute names) must be surrounded by double quotation marks. " can not exist in the middle of a string surrounded by double quotation marks.
8. * is a preserved key character
 - It is a pre-fix of node or edge section headers.
 - It is used as a delimiter between an attribute name and the corresponding data type.
 - It also represents any null/unknown value
 - * in strings surrounded by double quotation marks should be treated as a regular character.
9. All integers must be represented by a positive or negative number with no decimal.
10. All floating point numbers must include a decimal. -1.23e5 is acceptable notation, as are values such as 102.5
11. id, source and target MUST be integers. label MUST be strings.
12. Arbitrary number of spaces and/or tabs can exist between two columns.
13. After each section header, the meaning of each column MUST be specified in the next line with the format of attribute_name*data_type. Any comment line or empty is not allowed between the line of section header and the attribute line.
14. Attribute names CANNOT contain spaces and/or * in the middle and cannot be quoted. The values of data_type can ONLY be int, string, and float. attribute_name*data_type MUST be lower case.

Example2.nwb

```

*Nodes          4
id*int          label*string          weight*float          node_type*string
1              "Joe Ann"              0.66                 "author"
2              "John Smith"           0                    "author"
3              "Bio Today"            0.78                 "paper"
4              "Physics Tomorrow"     1.0                  "paper"
*DirectedEdges 3
source*int      target*int          weight*int          edge_type*string
1              3                  1                    "wrote"
4              3                  15                   "paper-citation"
2              3                  1                    "wrote"

```

15. All null/unknown values (data type could be string, or integer or floating point) are always represented by *. An empty string is "" (two double quotations marks)
16. Each comment line must start with # and comment lines must start from the beginning of a line and can not append to any node/edge section headers, attribute lines or node/edge lists.

Example3.nwb

```

*Nodes          4
id*int          label*string          weight*int          node_type*string
#the following node has an unknown string value.
1              *                    0                    "author"
2              "John Smith"           0                    "author"
3              "Bio Today"            8                    "paper"
#the following node has an unknown integer value.
4              "Physics Tomorrow"     *                    "paper"
*DirectedEdges  3
source*int      target*int          weight*float          edge_type*string
1              3                  0.66                 "wrote"
4              3                  0.78                 "paper-citation"
2              3                  1.0                  "wrote"

```

17. Separate node and edge attribute sections will continue to NOT be supported.

Acknowledgements

NWB File Format is defined by the NWB Development Team