Cuba Salud
IV Convención
Internacional de Salud
17-21 de octubre, 2022

Algoritmos para la estructuración de datos de neurociencias bajo la especificación BIDS

José Carlos Mendoza Sosa[1]
Dr. C. Arturo Orellana García [1]
Dr. C. Eduardo Aubert Vazquez[2]

[1] Centro de Informática Médica, Universidad de las Ciencias Informáticas. La Habana, Cuba, josecms@estudiantes.uci.cu, aorellana@uci.cu
[2] Centro de Neurociencias de Cuba. La Habana, Cuba

Resumen: El propósito principal de las Neurociencias es entender cómo el encéfalo produce la marcada individualidad de la acción humana. Los actuales estudiosos del cerebro, saben que para comprenderlo hay que derrumbar las barreras de las disciplinas tradicionales para mencionar apenas algunas de las áreas que han sido creadas, para caracterizar los métodos de estudio. Dentro de esta ciencia se puede encontrar los neurodatos, los cuáles son el conjunto de información relativa a la actividad cerebral obtenida mediante el empleo de neurotecnologías **(1)**. Los datos de neuroimagen son heterogéneos en su forma general, pueden comprender información en muchos formatos diferentes, que hacen que el intercambio de información se dificulte y se retrasen en ocasiones los resultados que podrían obtenerse de estudios científicos. Brain Imaging Data Structure (BIDS) es un estándar que trata de abordar este problema, pero existen aún mucha información que está fuera de el estándar BIDS y se utilizan herramientas, denominadas convertidores para estandarizar las bases de datos que no siguen las normativas de BIDS. El presente trabajo aborda la estandarización de neurodatos, se hace un análisis de los convertidores y algoritmos, que se emplean en la actualidad para llevar a cabo la estandarización, con el objetivo de resumir sus características principales y que sea de ayuda para los investigadores al momento de seleccionar uno de ellos. Se emplean métodos de análisis y síntesis para comprender los elementos que los componen, las funciones que cumplen y su importancia. Se obtiene una síntesis de las características de los convertidores, sus características comunes y se describe el escenario de los convertidores actualmente.

Palabras clave: BIDS, neuroimagen, conversión, neuroinformática, neurotecnologías

I. INTRODUCCIÓN

El término Neurociencias aparece en el diccionario como: “Ciencia que se ocupa del sistema nervioso o de cada uno de sus diversos aspectos y funciones especializadas”(2). Su empleo actual corresponde a la necesidad de integrar las contribuciones de las diversas áreas de la investigación científica y de las ciencias clínicas para la comprensión del funcionamiento del sistema nervioso(1).

El propósito principal de las Neurociencias es entender cómo el encéfalo produce la marcada individualidad de la acción humana. Los actuales estudiosos del cerebro, saben que para comprenderlo hay que derrumbar las barreras de las disciplinas tradicionales para mencionar apenas algunas de las áreas que han sido creadas, para caracterizar los métodos de estudio. Esta tendencia queda muy evidente en las obras científicas recientes las cuales tratan de las funciones más complejas de este órgano, como las emociones y la consciencia, apoyándose en los principales conceptos provenientes de las diversas disciplinas (2).

Dentro de esta ciencia se puede encontrar los neurodatos, los cuáles son el conjunto de información relativa a la actividad cerebral obtenida mediante el empleo de neurotecnologías avanzadas tales como FMRI (imágenes por resonancia magnética funcional), EEG (Electroencefalograma), PET (Tomografía de emisión de positrones), entre otras (1).

Las neurotecnologías son un conjunto de herramientas que sirven para analizar e influir sobre el sistema nervioso del ser humano, especialmente sobre el cerebro (3). Tecnologías que ayudan en el diagnóstico e intervención de diversas patologías psicológicas, emocionales y físicas, permitiendo tener un conocimiento más profundo de los síntomas y padecimientos del paciente.

Los datos de neuroimagen son muy heterogéneos. en su forma general, puede comprender información en muchos formatos diferentes, que contienen desde cantidades escalares simples hasta cadenas y matrices de datos multidimensionales. la gran variedad de protocolos, nomenclaturas e instrumentos existentes hacen que el intercambio de datos sea un desafío exigente en el campo. Abordar este problema es crucial para facilitar colaboraciones entre colegas y centros, así como para mejorar la reproducibilidad y transparencia de los resultados. Además, se convierte en un elemento fundamental aspecto organizativo para organizar grandes bases de datos basadas en numerosos sujetos, cada uno de ellos escaneado con múltiples instrumentos de imagen que proporcionan información complementaria de la estructura y función del cerebro (4).

Los experimentos de neuroimagen dan como resultados datos complicados que se pueden organizar de muchas maneras diferentes. Hasta el momento no hay consenso sobre cómo organizar y compartir los datos obtenidos en los experimentos de neuroimagen. Incluso dos investigadores que trabajan en el mismo laboratorio pueden optar por organizar sus datos de forma diferente. La falta de consenso (o de un estándar) conduce a malentendidos y a la pérdida de tiempo reorganizando datos o escribiendo guiones esperando cierta estructura. Con la Brain Imaging Data Structure (BIDS), se describe una forma simple y fácil de adoptar para organizar los datos de neuroimagen y de comportamiento(5).

Hasta la fecha no ha habido consenso sobre cómo organizar y compartir estos datos, lo que conduce a malentendidos y pérdida de tiempo. Algunos de estos formatos pueden ser: dcm o también llamados archivos de imágenes DICOM siglas de *Digital Imaging and Communications in Medicine* (6), el formato mnc (6), nifti (*Neuroimaging Informations Technology Initiative*) (6), *Analyze* (6), edf (7,8), PAR/REC (9), ECAT7 (10), BIDS(*Brain Imaging Data Structure*) (5).

Basado en el volumen de publicación, cada año se escanean decenas de miles de sujetos con fines de investigación. Cada estudio da como resultado datos complejos que involucran muchos archivos en diferentes formatos que van desde archivos de texto simples hasta archivos multidimensionales datos de imagen, que se pueden organizar de muchas maneras diferentes. De hecho, un estudio generalmente se compone de múltiples protocolos de imágenes y, a menudo, múltiples grupos de sujetos. La falta de consenso conduce a malentendidos y tiempo desperdiciado en reorganizar datos o reescribir scripts que esperan formatos de archivo particulares y organización, así como una posible causa de errores.(11)

Si bien no existe un acuerdo oficial sobre que formato o estándar usar, Brain Imaging Data Structure (BIDS)es en la actualidad un estándar en apogeo y muchos grupos investigativos lo utilizan para intercambio. Es válido suponer que en función de establecer intercambios científicos fructuosos se desea trabajar desde estándares establecidos como lo es o apunta a serlo el BIDS.

Surge entonces la necesidad de generar información que esté estructurada según las normativas de BIDS, la información nueva sería relativamente fácil, sin embargo, no sería práctico repetir todos los estudios que no seguían las normativas de BIDS. Es donde los convertidores ganan su espacio generando datasets válidos para BIDS a partir de datasets que estaban en otros formatos: como DICOM, MINC, entre otros.

El presente trabajo tiene como objetivo analizar los convertidores a BIDS más representativos en la literatura especializada, para identificar sus características funcionales más importantes y detectar las limitantes que poseen para estandarizar bases de datos de neurociencias.

II. MÉTODO

Para la ejecución de la presente investigación se sigue una estrategia explicativa y se emplearon: los métodos: Análisis y Síntesis: análisis sobre los métodos de estandarización de imágenes existentes, determinando los elementos que los componen, las funciones que cumplen y su importancia. Determinar que parámetros se deben tener en cuenta para la selección del método a utilizar. Mediante el método comparativo: establecer la comparación entre los diferentes métodos de estandarización.

III. RESULTADOS

A. Dcm2bids

dcm2bids es un proyecto centrado en la comunidad. Su objetivo es ser una herramienta amigable y fácil de usar para convertir sus DICOMs.(12) Requiere dcm2niix previamente instalado y luego se puede usar el gestor de paquetes de Python (pip) para instalarlo. Es un convertidor que se ejecuta en la terminal de comandos. Tiene en su haber varias herramientas que ayudan al flujo de trabajo de la conversión al estándar BIDS como *dcm2bids_scaffold* para estructurar directorios y *dcm2bids_helper* para brindar ayuda e información sobre los comandos del convertidor y sus banderas. La conversión requiere además de un archivo de configuración. El archivo de configuración requiere que se cree una configuración por cada adquisición, permite que apliquemos una configuración a varios archivos por expresiones regulares en una llave denominada *criteria.*

```
{
  "descriptions": [
    {
      "dataType": "func",
      "modalityLabel": "bold",
      "customLabels": "task-rest",
      "criteria": {
        "SeriesDescription": "Axial EPI-FMRI (Interleaved I to S)*"
      "sidecarChanges": {
        "TaskName": "rest"
      }
      }
    }
  ]
}
```

Figura 1 Archivo de configuración (12)

En la Fig.1 se indica que a cada adquisición que contenga "*Axial EPI-FMRI* (*Interleaved I to S*)" en el campo *SeriesDescription* se le aplique esa configuración. Es preciso aclarar que este convertidor permite la conversión de una sesión de neuroimágenes de un solo sujeto a la vez.(4) Al momento de escribir este artículo dcm2bids se mantiene actualizado (GitHub).

B. Bidskit

Bidskit intenta realizar un seguimiento de la Especificación BIDS lo más fielmente posible. Requiere de una instalación previa de dcm2niix y luego se puede usar el gestor de paquetes de *Python* (pip) para instalarlo. Para comenzar, debe crear una carpeta de dataset con un nombre semidescriptivo (p. ej., *learning_pilot_2019*) con una subcarpeta *sourcedata*/ que contenga sus datos DICOM sin procesar, organizados por tema o por tema y sesión. Un árbol de directorio DICOM típico podría parecerse al siguiente, donde Cc0001, Cc0002 son ID de sujeto y primero, "segundo" son nombres de sesión.

```
learning_pilot_2019/
├── sourcedata
|   ├── Cc0001
|   |   ├── first
|   |   |   └── [DICOM Images]
|   |   └── second
|   |       └── [DICOM Images]
|   └── Cc0002
|   |   ├── first
|   |   |   └── [DICOM Images]
|   |   └── second
|   |       └── [DICOM Images]
...
```

Figura 2 Árbol de directorios (13)

Es un convertidor que ejecuta sus tareas y comandos en la terminal de comandos. Bidskit se caracteriza por sus comandos muy intuitivos y poco complicados para el usuario final. Para llevar a cabo una conversión se debe correr bidskit en dos ocasiones la primera es para estructurar adecuadamente el árbol de directorios (en caso de ser preciso) y la segunda donde se ejecuta la conversión al estándar BIDS.

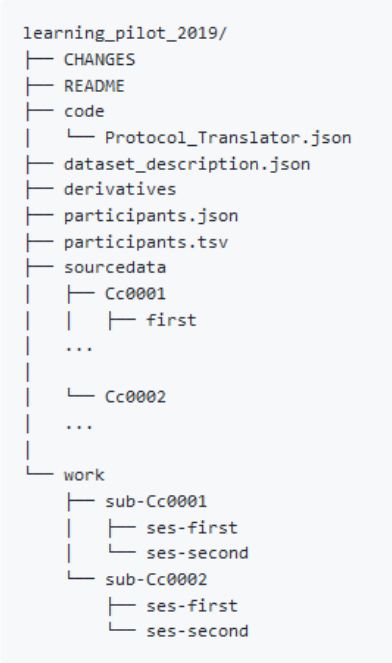

Figura 3 Árbol de directorios luego de la primera ejecución de bidskit (13)

El árbol de directorios luego de haber ejecutado bidskit por primera vez, se crean varios archivos y directorios, entre ellos el *Protocol_Translator.json* que bidskit emplea para ejecutar la segunda ejecución y tener una conversión a BIDS exitosa. Bidskit luego de haber sido modificado el archivo de configuración correctamente, está preparado para realizar la conversión. Al momento de escribir el artículo bidskit se mantiene actualizado (GitHub).

C. Bidscoin

BIDScoin es una aplicación de *Python* de código abierto y fácil de usar que convierte ("*coins*") datasets de neuroimagen (sin procesar) a nivel de fuente en conjuntos de datos nifti / json / tsv que se organizan de acuerdo con el estándar *Brain Imaging Data Structure* (BIDS). En lugar de depender de una lógica programática compleja para la identificación del tipo de datos de origen, *BIDScoin* utiliza un enfoque de mapeo para descubrir los diferentes tipos de datos de origen en su repositorio y convertirlos en tipos de datos de BIDS. Las diferentes ejecuciones de datos de origen se identifican de forma única por sus propiedades del sistema de archivos (p. ej., nombre o tamaño del archivo) y por sus atributos (p. ej., *ProtocolName* del encabezado DICOM). La información de mapeo se puede especificar previamente (por ejemplo, por sitio), lo que permite a *BIDScoin* hacer primeras sugerencias inteligentes sobre cómo clasificar y convertir los datos. Si bien este procedimiento de línea de comandos explota toda la información disponible en el disco, *BIDScoin* presenta una interfaz gráfica de usuario (GUI) para que los investigadores verifiquen y editen estas asignaciones, aportando el conocimiento faltante que a menudo existe solo en sus cabezas.(14)

Requiere tener instalado dcm2niix y el gestor de paquetes de Python (pip) se instalaría sin problemas.

BIDScoin requiere que el repositorio de datos de origen esté organizado de acuerdo con una estructura de sujeto/[sesión]/datos (la subcarpeta de la sesión es opcional). La carpeta de datos se puede estructurar de varias maneras, como se ilustra en los siguientes ejemplos:

1. Organización '*seriesfolder'*.
2. Organización 'DICOMDIR'.
3. Organización plana de DICOM.
4. Organización PAR/REC.

Bidscoin ofrece dos herramientas para el preprocesamiento *dicomsort* y *rawmapper*. *Dicomsort* es capaz de convertir cualquiera de las 4 organizaciones antes mencionadas a la que es requerida por *BIDScoin*. *rawmapper* puede mostrarle una descripción general (*map*) de todos los valores de los atributos DICOM de interés en su conjunto de datos y, opcionalmente, se utiliza para cambiar el nombre de las subcarpetas de datos de origen. La última opción puede ser útil, ej. sí ingresó manualmente identificadores de materias como [Información adicional] en la consola del escáner y desea utilizarlos para cambiar el nombre de sus carpetas de sujetos.

Luego de la preparación de los directorios de los datasets, viene el trabajo con el editor donde el usuario asegura los valores que son extraídos automáticamente por el convertidor, valores que utiliza un archivo de configuración que le servirá luego para ejecutar la conversión automática del dataset. Este proceso es mejor descrito en la documentación oficial del convertidor.(15)

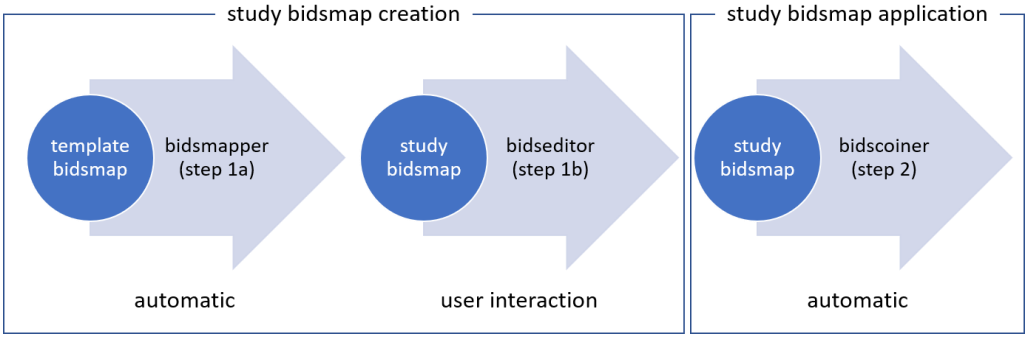

Figura 4 Flujo de trabajo de BIDScoin (15)

Al momento de escribir el artículo *BIDScoin* se mantiene actualizado (GitHub).

D. Heudiconv

Heudiconv es un convertidor DICOM flexible para organizar datos de imágenes cerebrales en diseños de directorios estructurados.

- Permite diseños de directorios flexibles y esquemas de nombres a través de implementaciones heurísticas personalizables
- Solo convierte los DICOM necesarios, no todo en un directorio
- Puede mantener enlaces a archivos DICOM en el diseño del participante
- Usando dcm2niix por debajo de su implementación, es rápido.
- Puede rastrear la procedencia de la conversión de DICOM a NIfTI en formato W3C PROV
- Proporciona asistencia para convertir a BIDS.
- Se integra con *DataLad* para colocar los datos convertidos y originales bajo el control de versiones de git/git-annex, mientras anota automáticamente los archivos con información confidencial (p. ej., anatómicos no desfigurados, etc.)(16)

Similar a otros convertidores, heudiconv necesita ejecutar una vez antes de realizar la conversión. Durante esta conversión crea carpetas donde se almacenarán los datos convertidos y otras que le serán útil para la conversión, como el archivo de heurística, por defecto heuristic.py pero, puede tener otro nombre, incluso, asumir varios y cada uno se ajusta a un *dataset* específico.

El archivo de configuración de heudiconv es intuitivo siempre que los datos tengan correctos sus metadatos, es suficiente crear llaves en diccionario para cada modalidad de adquisición empleadas en el dataset. Heudiconv tomará todos los archivos de esa modalidad y les dará tratamiento de acuerdo a su modalidad y arrojando un dataset convertido con solo conocer las modalidades o secuencias empleadas en el dataset. Al momento de escribir el artículo *BIDScoin* se mantiene actualizado (GitHub).

IV. CONCLUSIONES

Existe un problema de heterogeneidad en los datos de neurociencias que provocan retrasos en las investigaciones de neurociencias.

El estándar BIDS trata de abordar el problema de la heterogeneidad en los neurodatos.

Los convertidores analizados usan un archivo de configuración para ejecutar la conversión. Requieren además de una organización previa de la estructura del árbol de directorios. dcm2bids y bidscoin ofrecen herramientas para la estructuración del árbol de directorios.

Los convertidores analizados requieren modificaciones para cubrir las necesidades de los investigadores.

AGRADECIMIENTOS

La investigación que da origen a los resultados presentados en la presente publicación recibió fondos de la Oficina de Gestión de Fondos y Proyectos Internacionales bajo el código PN305LH013-038.

REFERENCIAS

1. El riesgo de los neurodatos: nuevas perspectivas sobre su utilización - {The} {Technolawgist}.
2. neuroaprende. {NEUROCIENCIAS}: término, origen y actualidad. neuroaprende. 2018.
3. Kobayashi M, Pascual-Leone A. Transcranial magnetic stimulation in neurology. The Lancet Neurology. 2003;2(3):145–56.
4. Lopez-Novoa U, Charron C, Evans J, Beltrachini L. The BIDS Toolbox: A web Service to Manage Brain Imaging Datasets. In: 2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI). 2019. p. 378–82.
5. Gorgolewski KJ, Auer T, Calhoun VD, Craddock RC, Das S, Duff EP, et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. Scientific Data. 2016 Jun;3.
6. Larobina M, Murino L. Medical Image File Formats. J Digit Imaging. 2014 Apr;27(2):200–6.

7. EDF Extensión de archivo - ¿Qué es .edf y cómo abrir? - ReviverSoft [Internet]. [cited 2021 Jun 10]. Available from: https://www.reviversoft.com/es/file-extensions/edf
8. KEMP B. European data format (EDF): current availability and additional applications. J Sleep Res. 2002;11(Suppl 1):120.
9. Tellado Vazquez B. Tesis Madrid--Herramienta basada en Matlab para larealización y análisis de estudiosestructurales y funcionales a partir deimágenes de resonancia magnética. Universidad Carlos III de Madrid;
10. Image Formats [Internet]. [cited 2021 Jun 11]. Available from: http://www.nf.mpg.de/vinci3/doc/image-formats.html
11. About DICOM: Overview [Internet]. [cited 2022 Jun 10]. Available from: https://www.dicomstandard.org/about-home
12. Bedetti C, arnaudbore, Guay S, Carlin J, Nick, Dastous A, et al. UNFmontreal/Dcm2Bids: 2.1.8 [Internet]. Zenodo; 2022 [cited 2022 Jul 13]. Available from: https://zenodo.org/record/6658099
13. Tyszka M. BIDSKIT [Internet]. 2022 [cited 2022 Jun 13]. Available from: https://github.com/jmtyszka/bidskit
14. Zwiers MP, Moia S, Oostenveld R. BIDScoin: A User-Friendly Application to Convert Source Data to Brain Imaging Data Structure. Frontiers in Neuroinformatics. 2022 Jan;15:65.
15. The BIDScoin workflow — BIDScoin 3.7.2 documentation [Internet]. [cited 2022 Jul 13]. Available from: https://bidscoin.readthedocs.io/en/stable/workflow.html
16. HeuDiConv — heudiconv 0.5.4 documentation [Internet]. [cited 2022 Jul 13]. Available from: https://heudiconv.readthedocs.io/en/latest/