



## **BrainSSys: Plataforma para el almacenamiento, estructuración, estandarización y análisis de datos de neurociencias**

Dr.C. Arturo Orellana García<sup>1</sup>

Jesús E. Fuentes González<sup>1</sup>

Dr.Cs. Pedro Valdés Sosa<sup>2</sup>

Dra.C. Lídice Galán García<sup>2</sup>

Dr.C Eduardo Martínez Montes<sup>2</sup>

<sup>1</sup> Centro de Informática Médica, Universidad de las Ciencias Informáticas, La Habana, Cuba, contacto: aorellana@uci.cu

<sup>2</sup> Centro de Neurociencias de Cuba, BioCubaFarma, La Habana, Cuba, contacto: pedro.valdes@neuroinformatics-collaboratory.org

**Resumen:** Existe un interés creciente y aplicativo a nivel internacional por las neurotecnologías en áreas de diagnóstico médico, análisis y procesamiento de datos de neurociencias. El volumen de datos que se genera en estos estudios aumenta exponencialmente con la adquisición de nuevas tecnologías y equipos biomédicos para obtenerlos. Esto hace que la gestión de los datos sea más difícil y su heterogeneidad más compleja para el procesamiento y análisis. El trabajo tiene como objetivo presentar el proyecto de estructuración, estandarización y procesamiento de datos de neurociencias BrainSSys.

Se consultó y analizó la documentación científico técnica disponible de plataformas con similar propósito para identificar las mejores prácticas, estándares y formas de gestión de los datos. Para el desarrollo se utilizan tanto bases de datos relacionales como no relacionales, donde prevalece el uso de MongoDB. Los lenguajes de programación que predominan son Python y Java. Para el despliegue y orquestación de microservicios se aplica Docker; mientras que Apache Kafka es usado como almacén de eventos y la comunicación interna de la aplicación.

La plataforma BrainSSys consta de 3 subsistemas principales: DataBrain, PixelBrain y SyncBrain. Un repositorio para la estandarización y tratamiento de los datos de neurociencias. Herramientas de procesamiento, análisis y visualización de neurodatos; por último, un sincronizador de datos entre plataformas.

BrainSSys propiciará a CNEURO contar con herramientas de gestión propias, que garanticen su soporte y evolución funcional teniendo en cuenta sus necesidades tecnológicas. Propiciará la estandarización y análisis de los datos y brindará herramientas que faciliten el desarrollo de pipelines de investigaciones.

**Palabras clave:** datos de neurociencias, estandarización, procesamiento, software.

## I. INTRODUCCIÓN

Las neurociencias son las ciencias multidisciplinarias que analizan el sistema nervioso para comprender las bases biológicas del comportamiento del cerebro (estructura, función, desarrollo ontogenético y filogenético, bioquímica, farmacología y patología). Los neuroanatomistas estudiaron la forma del cerebro, su estructura celular y sus circuitos; los neuroquímicos estudian la composición química del cerebro, sus lípidos proteínas; los neurofisiólogos estudian las propiedades bioeléctricas del cerebro; y los psicólogos y neuropsicólogos investigan la organización y sustratos neuronales del comportamiento y la cognición. (1)

En esta área se utilizan varias técnicas imagenológicas para el diagnóstico de enfermedades y la asistencia en intervenciones quirúrgicas con el objetivo de determinar la ubicación y dimensión exacta de masas y tumores cerebrales, así como estudios y seguimientos a distintas patologías. El desarrollo tecnológico propicia la aparición de nuevos equipos de adquisición de imágenes y señales que abren el margen de oportunidades hacia investigaciones más pertinentes y específicas para el análisis del cerebro. (2,3) Los equipos de adquisición generan múltiples formatos de salidas lo cual dificulta la estandarización en el uso y aplicación de algoritmos para el análisis de neurodatos y el comportamiento de patologías en datos longitudinales. Lo anterior fue enfatizado como un problema internacional por la comunidad científica en el Global Brain Consortium GBC 2020, desarrollado en el Balneario de Varadero, Cuba.

Existe un interés creciente y aplicativo a nivel internacional de las tecnologías de neuroimágenes en áreas de diagnóstico médico, análisis y documentación histórica, a lo que estas técnicas se refieren. El aumento en esta área de investigación puede ser observada por el largo número de documentos científicos publicados en blogs y revistas científicas desde el año 2000. Esto puede ser atribuido por el incremento de dispositivos para el diagnóstico por la comunidad médica.

Cada equipo médico de adquisición de estudios del cerebro genera ficheros en formatos diferentes como dcm, edf, nii, plg, mnc, entre otros. Esta variedad de extensiones dificulta el análisis y la estandarización de las fuentes de datos aptas para el proceso de investigación. Los estudios de neuroimagen dan como resultado datos complicados que se pueden organizar de muchas formas diferentes. (4-7)

Hasta hace poco no existía consenso sobre cómo organizar y compartir datos obtenidos en estudios de neurodatos. Incluso dos investigadores que trabajan en el mismo laboratorio podían optar por organizar sus datos de manera diferente. La falta de consenso (o un estándar) conducía a malentendidos y pérdida de tiempo en reorganizar datos o reescribir guiones que esperan cierta estructura. A partir del 2018 comenzó a generalizarse el estándar BIDS (Brain Imaging Data Structure), el cual solucionó parcialmente el problema existente. (8)

La neuroinformática es un área de la ciencia cuyo objetivo es integrar los datos de la neurociencia y desarrollar modernas herramientas informáticas para aumentar la comprensión de las funciones del sistema nervioso en la salud y la enfermedad. Las herramientas neuroinformáticas incluyen, entre otras, bases de datos para almacenar y compartir datos, repositorios para gestionar documentos y código fuente, y herramientas de software para analizar, modelar y simular señales e imágenes. (9)

Esta ciencia combina la neurociencia y la informática para desarrollar y aplicar herramientas y enfoques avanzados esenciales para comprender la estructura y la función del cerebro. También se considera que la neuroinformática engloba la bioinformática más tradicional y la investigación de biología de sistemas computacional recientemente establecida para la neurociencia. El campo de la neuroinformática ha crecido rápidamente hasta implicar el desarrollo de modernas herramientas basadas en las tecnolo-

gías de la información y la comunicación (TIC). Los diferentes gremios de investigación en el mundo desarrollaron convertidores a partir de algoritmos computacionales para llevar de formatos específicos al estándar BIDS, sin embargo, cuando los Datasets son multimodales estos no funcionan. (5)

El aumento de los volúmenes de datos que se generan, por ejemplo, entre 1999 y 2010 el número de imágenes de una Tomografía Digital pasó de 82 a 679 por examen, las Resonancias Magnéticas pasaron de 164 a 570 imágenes. A esto se suma que el desarrollo de nuevos equipos multifunción, marca un avance acelerado para potenciar soluciones en la detección de enfermedades y afecciones, sobre todo en el cerebro. (10)

El programa Nacional de Neurociencias y Neurotecnologías ha propiciado que la Universidad de las Ciencias Informáticas y el Centro de Neurociencias de Cuba colaboren para desarrollar soluciones a las problemáticas existentes sobre el manejo, gestión y procesamiento de datos de neurociencias. El presente trabajo tiene como objetivo presentar el proyecto de estructuración, estandarización y procesamiento de datos de neurociencias BrainSSys.

## II. MÉTODO

Para el desarrollo de la propuesta se consultó y analizó la documentación científico técnica disponible de plataformas con similar propósito para identificar las mejores prácticas, estándares y formas de gestión de los datos. Para el desarrollo se utilizan tanto bases de datos relacionales como no relacionales, donde prevalece el uso de MongoDB. Los lenguajes de programación que predominan son Python y Java. Para el despliegue y orquestación de microservicios se aplica Docker; mientras que Apache Kafka es usado como almacén de eventos y la comunicación interna de la aplicación.

## III. RESULTADOS

La Plataforma BrainSSys es un ecosistema de software neurocientífico orientado al almacenamiento, gestión, visualización y procesamiento de datos provenientes de estudios del cerebro humano. Desarrollada con tecnologías emergentes y escalables con múltiples ventajas para su aplicación en entornos heterogéneos. Consta de 3 subsistemas principales: DataBrain, PixelBrain y SyncBrain. El primero es un repositorio para la estructuración, estandarización y tratamiento de los datos de neurociencias. El segundo está compuesto por herramientas de procesamiento, análisis y visualización de neurodatos; por su parte, SyncBrain propicia la sincronización y transferencia de datos con instituciones extranjeras que colaboran con CNEURO.

### A. Descripción técnica de la plataforma

La plataforma, está conformada por diferentes subsistemas interconectados a partir de una arquitectura base de microservicios.

**DataBrain:** Repositorio para el almacenamiento y gestión de datos de neurociencias, posee funcionalidades para la información estadística y toma de decisiones. Incorpora herramientas para el manejo, seguridad de los datos y la actualización de sus bases de datos. Posee API's de servicios a otros sistemas. Sus principales herramientas son:

- Arquitectura modular para el almacenamiento de bases de datos de neurociencias.

- Una herramienta para estructurar bases de datos y estandarizar datos de neurociencia.
- Herramientas para la gestión, control y seguimiento de bases de datos cerebrales.
- Una Herramienta para el tratamiento y limpieza de datos en Bases de Datos Cerebrales. Algoritmos para la clasificación y completamiento de datos y metadatos.
- Algoritmos para la compresión de bases de datos sin pérdida de información.

**PixelBrain:** Herramienta de visualización, análisis y procesamiento de datos de neurociencias. Compuesta por algoritmos y técnicas que propician la mejora de información diagnóstica y científica; y detección de elementos de interés investigativo en datos de neurociencias.

- Herramienta para la visualización de datos multimodales de neurociencias.
- Algoritmos para la fusión, procesamiento y clasificación de datos de neurociencia.
- Herramientas para el análisis e interpretación de conjuntos de datos.

**SyncBrain:** Herramienta para la conexión y transmisión bidireccional de datos y bases de datos con plataformas internacionales. Permite la regulación de la transmisión en función de la velocidad y carga de red.

- Componente para establecer una conexión segura con las plataformas CBRAIN y LORIS.
- Herramienta para la transmisión bidireccional de datos comprimidos a las plataformas CBRAIN y LORIS.
- Herramienta informática para la sincronización con fuentes de datos cubanas.

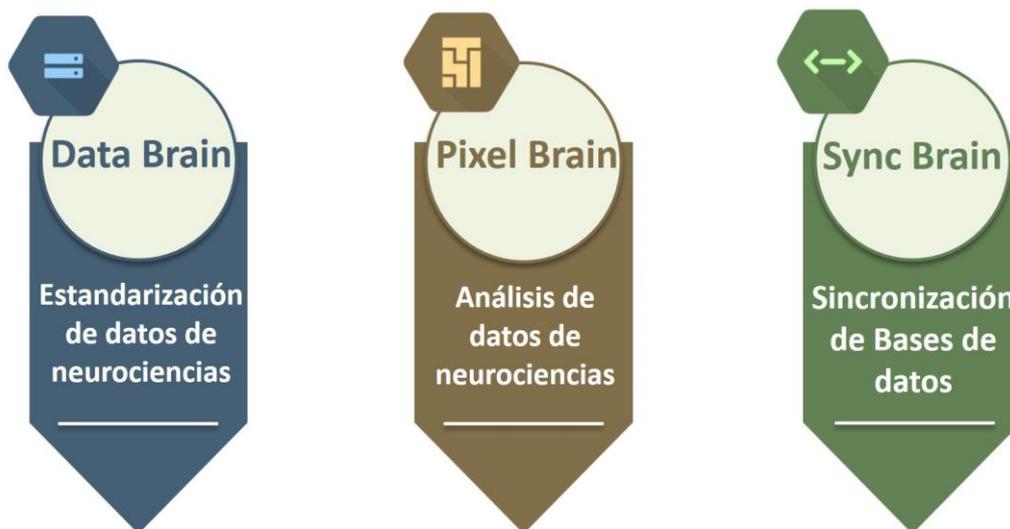


Figura 1. Subsistemas principales de la Plataforma BrainSSys. Fuente: los autores.

Además, cuenta con **ToolsBrain:** Conjunto de herramientas para el soporte y gestión adicional de datos de neurociencias. Propician la transformación y estandarización de datos de neurociencias, la gestión de colas de procesamiento en HPC, la modelación dinámica de Pipelines de datos, algoritmos de corre-gistro y fusión de datos espacio-temporales multimodales.

La Figura 2 muestra el diagrama de despliegue de la infraestructura de BrainSSys. Presenta tres espacios de nombre que agrupan microservicios según su responsabilidad y las tecnologías asociadas. Utiliza como orquestador de microservicios las tecnologías de Docker y un Api Gateway para hacer balances de cargas.

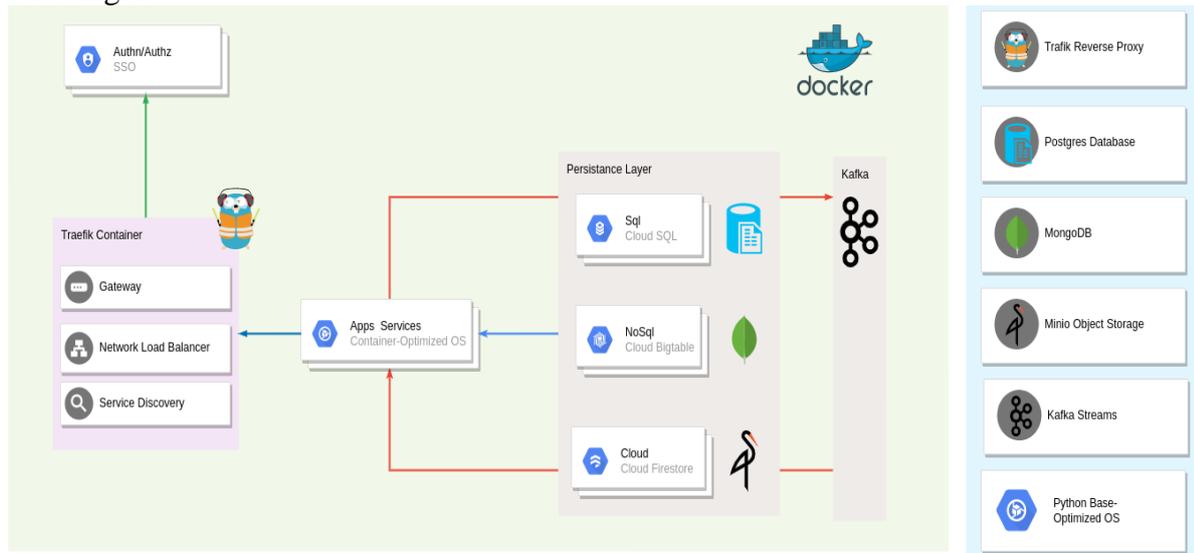


Figura 2. Diagrama de Despliegue de la infraestructura de BrainSSys. Fuente: los autores.

La figura 3 es una Representación de las capas de abstracción de la arquitectura del microservicio lógico del repositorio DataBrain y del microservicio de gestión de tuberías de ejecución ToolSet. El primero refina los datos pasando por varios procesos bottom up de extracción, transformación y carga. El segundo expone los mecanismos de gestión y ejecución de tuberías, así como sus metadatos asociados.

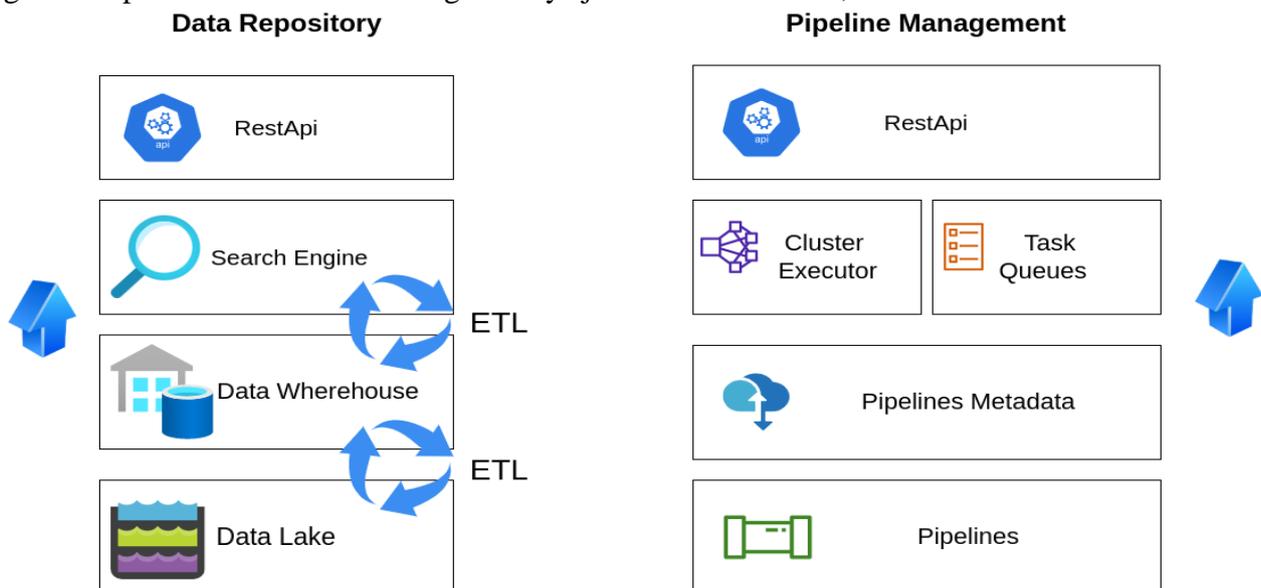


Figura 3. Capa de abstracción de arquitectura de los microservicios DataBrain y ToolsSet. Fuente: los autores.

Por su parte, la figura 4 representa la Arquitectura interna hexagonal basada en *Domain Driven Design* (DDD) o Diseño Dirigido por Dominio para el desarrollo de microservicios. Presenta tres capas principales, Capa de Dominio, Capa de casos de uso y Capa de aplicación. Su principal característica es que permite hacer inversiones de dependencias sobre los componentes técnicos que no se relacionan con el negocio manteniendo este invariable y capaz de evolucionar a nuevas tecnologías con facilidad. La infraestructura no se concibe como una capa interna sino como la capa externa del microservicio.

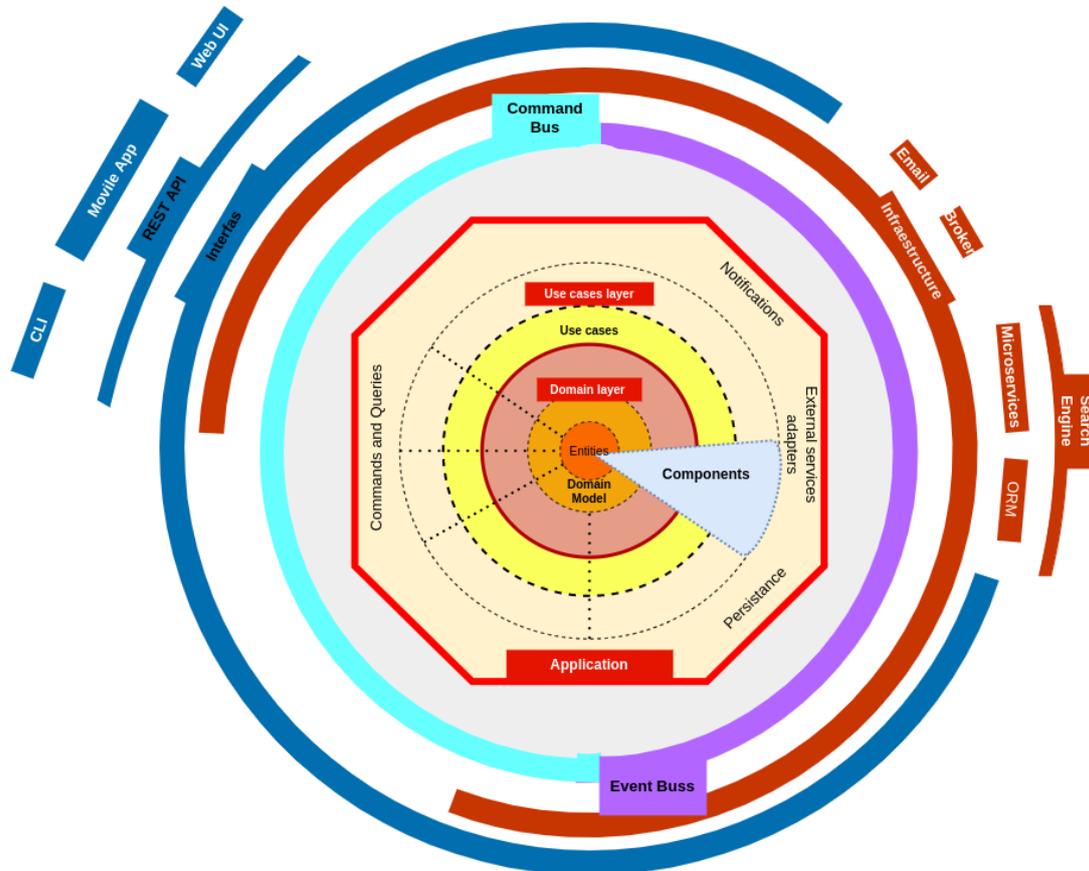


Figura 3. Arquitectura interna hexagonal basada en Domain Driven Design. Fuente: los autores.

### B. Impactos esperados

El Proyecto está acorde a la Política Económica y Social de Cuba y la política de Ciencia, Tecnología, Innovación y Medio Ambiente. Así mismo, responde a lineamientos de la Política Social en el sector de Salud con vistas a sostener y desarrollar los resultados alcanzados en el campo de la industria del software y la informatización de los procesos fundamentales en las instituciones de salud y los centros de investigación médica y biomédica. Fomentará la innovación científica y tecnológica con impacto en la salud de la sociedad cubana y así mismo la introducción de resultados al desarrollo de la medicina del país.

## IV. CONCLUSIONES

La plataforma podrá convertir y estandarizar conjuntos de datos para que los utilicen los investigadores en neurociencia. De esta forma, apoyar la toma de decisiones y nuevas investigaciones en el Centro Nacional de Neurociencias de Cuba.

Promoverá la innovación científica y tecnológica con impacto en la salud de la sociedad cubana y así mismo la introducción de resultados al desarrollo de la medicina en el país. Permitirá reducir errores en el procesamiento de datos, optimizar tiempos de ejecución, administrar y estandarizar bases de datos para reducir sesgos en los resultados de la investigación.

El resultado permitirá la creación de software nacional para la informatización de un grupo de procesos investigativo de alto nivel internacional

## 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. Pernet CR, Appelhoff S, Gorgolewski KJ, Flandin G, Phillips C, Delorme A, Oostenveld R. EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. *Scientific Data*. 2019 Jun 25;6(1):1-5.
2. Gorgolewski KJ, Auer T, Calhoun VD, Craddock RC, Das S, Duff EP, Flandin G, Ghosh SS, Glatard T, Halchenko YO, Handwerker DA. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific data*. 2016 Jun 21;3(1):1-9.
3. Niso G, Gorgolewski KJ, Bock E, Brooks TL, Flandin G, Gramfort A, Henson RN, Jas M, Litvak V, Moreau JT, Oostenveld R. MEG-BIDS, the brain imaging data structure extended to magnetoencephalography. *Scientific data*. 2018 Jun 19;5(1):1-5.
4. Singh R, Srivastava R, Prakash O, Khare A. Multimodal medical image fusion in dual tree complex wavelet transform domain using maximum and average fusion rules. *Journal of Medical Imaging and Health Informatics*. 2012 Jun 1;2(2):168-73.
5. Brusko, G. D., Gregory, B., and Wang, M. Y. "Big Data in the Clinical Neurosciences." *Machine Learning in Clinical Neuroscience*. Springer, Cham, 2022. 271-276.
6. Ganapathy, K., Shabbir, S. A., and Aldilas, A. N. "Artificial intelligence in neurosciences: A clinician's perspective." *Neurology India* 66.4 (2018): 934.
7. Forero, D. A., et al. "Scientific productivity in neurosciences in Latin America: a scientometrics perspective." *International Journal of Neuroscience* 130.4 (2020): 398-406.
8. Roehri, N., et al. "Data transfer, collection and organisation for multi-centric clinical studies." *Workshop GEANT: Data Management and Sharing in Neuroinformatics*. 2019.
9. Niso, G., et al. "MEG-BIDS, the brain imaging data structure extended to magnetoencephalography." *Scientific data* 5.1 (2018): 1-5.
10. Dong, L, et al. "WeBrain: A web-based braininformatics platform of computational ecosystem for EEG big data analysis." *NeuroImage* 245 (2021): 118713.