[[1]](#footnote-2)

Data Management of Electronic Medical Recordings on WebBioBank

AmirAli Farokhniaee\*, Lorenzo Rossi, Mattia Arlotti, and Sara Marceglia

*Abstract*— *Goal:* WebBioBank (WBB) is a web-based platform to collect and share anonymous patients’ biosignals with Parkinson’s disease (PD) undergoing clinical treatments such as deep brain stimulation. The goal of this study was to characteristically categorize WBB biosignals utilizing computational algorithms. *Methods:* We summarized the recordings information in a metadata file and developed Matlab scripts to format and classify the biosignals. *Results:* 561 signals from 57 PD patients were classified based on personal information and clinical status such as on/off medication and stimulation. The old .txt files of signals were modified and their IEEE JSON format were created. Additionally, signal gains were estimated if missing. *Conclusions:* The algorithms developed in this study can be used to organize any kind of big biosignal bank, where their manual edit would be a cumbersome task. The future plan is to embed online signal processing tools on WBB for scientific analysis and clinical investigations.

*Index Terms*—Enter up to 5 keywords in alphabetical order, separated by commas.

***Impact Statement*— The authors should include here an impact statement of no more than 30 words. The statement should summarize the main findings of the research work reported in the manuscript.**

# INTRODUCTION

F

or more than 15 years of advanced research, WebBioBank (WBB) collects biosignals coming from patients recorded in different condition and from different brain areas developed by Newronika … It has been established since X and has been used in X institutions/research centres.

We present a background of Webbiobank (WBB) and in the next session we address the new features and updates.

>> Data base and server structure of WBB. The WBB system architecture is explained in detail in [Rossi JBI 2014]

Languages and structure:

data banks server and client sides: For each patient there are questioners

>>> signals of 10 minutes, beta is computed every minute. Now on the app and WBB the signal is shown plus the beta power in a desired time period.

These are done for both left and right electrodes in each signal.

>> Implementation of signal processing tools such as spectrograms, filtering in different rhythms, calculation of phase-amplitude coupling etc.

**Updates on WBB, the modifications and new stuff**

The view of the signals and their features such as … are available to the authorized researchers and clinicians.

>> Modifications done: modifying the tabs and appearance, adding users and patients.

We went over modifying the data tables and case report form (CRF).

It is going to be implemented on the WBB patient's data (info). Also amount of medication etc. (Mention what is done new…)

>> We went over the device tab as well. This shows the device names and their properties. In the device list we have

the list of those devices that are paired to the patients, and which patients they are paired with.

Manufacturer medtronic and activaPS (or TM SC, not sure, check later).

>> In the patients list and each patient information, there are three tabs, questionaire, signals and devices.

Signals are the signals recorded from a specific area of the patient and devices are the attached devices to the patient.

Recommendations and alternative approaches, lessons learned, comparison with current systems

The matlab codes are applied on part of the WebBioBank (WBB) developed by Newronika S.p.A. This data set contains local field potential (LFP), Electroencephalographical (EEG), Electromyographical (EMG) and other types of recordings from patients with Parkinson's disease (PD) gathered over years from both male and female genders and different range of ages. The pateints were undergone deep brain stimulation (DBS) implantation surgery, by which the LFPs where recorded in different settings such as off/on DBS, off/on medication, rest/move etc.

# Materials and Methods

This section explains the methods including the Matlab script codes that are used to clean up and categorize the data on Newronika WebBioBank data base. The codes and their corresponding descriptions are available at a github repository: <https://github.com/aafarokh/WBB_NWK>. We have converted the existing text files for the signals in diverse formats (old files) to unique txt file format (new files), explained in section II. B, below. The codes can be used to clean and group any kind of biological and brain recording data bank, where big amount of data is stored and their manual selection and edit would be a cumbersome task.

## Metadata file

The first step in managing big amount of data can be creation of sheets that include a summary of each recording information. A unique pateint ID (PID) is given to each patient name. Each signal is labeled by a signal ID (SID) that is unique. At the end, each old file is named by combining PID and SID so that the new file name is in the format of PID*abcd*\_SID*vwxyz* where small letter are integers. The information of each recording and subject is gathered in a metadata file as a sheet, such as in Microsoft Office Excel format. Each row presents one of our signals with the corresponding filename PID*abcd*\_SID*vwxyz* and the columns gather the overall information of the subject and the recorded signal. The columns in the metadata file are explained in Table I. (provide figure)

## Converting old files to the new ones (FileAdopter.m)

The objective of this data management is to read the old files in the data base, which contain different format of saving the signals and convert them to a one-column .txt file (new files):

*"Fs= " z*

*v(1)*

*v(2)*

*v(3)*

*v(4)*

*.*

*.*

*.*

*v(end)*

where *z* is the sampling frequency (*Fs*) of the signal in Hz and *v* is the value of the signal at a given time sample. All signal files are in .txt format, either the old or the new versions. The files that contain different signals from different subjects in the old format could have different rows and columns such as:

*"Time" "31 Keyboard" "2 untitled" "1 untitled"*

*0.0000 0.000000*

*0.0004 0.000000 0.62805 0.88745*

*0.0008 0.000000 1.20544 0.90698*

*0.0012 0.000000 0.89966 0.61035*

The first column contains the time in most of the cases and the other rows contain the signals. The signals could be different types of the same subject, such as LFP, EMG or different channels of one LFP recording.

TABLE I

Metadata file columns description

|  |  |
| --- | --- |
| Column | Description |
| FileName | The new file name in the format PIDxxxx\_SIDxxxxx |
| PatientID | The unique ID for each patient |
| SignalID | The unique ID of each signal |
| SourceFile | The name of the old file |
| FileLocation | The location of the old file |
| SubjectType | Type of the subject in the study  e.g. patient or animal |
| SubjectName | Name of the subject |
| SubjectGender | The gender of the subject, M/F |
| SubjectDoB | Subject date of birth |
| SubjectCondition | The subject's pathology, e.g. PD |
| SignalType | The type of signal e.g. LFP, EEG |
| SignalContent | Voltage trace or spectral content |
| BodySite | The recording site on the subject body |
| Channel | The recording channels |
| Medication | Medication state ON/Off |
| Stimulation | Stimulation state ON/Off |
| SignalUnit | Unit of the recorded signal |
| Number\_Column\_SourceFile | This column is used for the data conversion code and mentions the column in the old file where the new file is constructed on |
| SignalFs | Signal sampling frequency (Hz) |
| SignalGain | Gain of the signal |
| EstimatedGain | The estimated gain in case the gain was not reported in the old file and corresponding data base |
| SignalBandWitdth | Bandwidth of the recorded signal |
| Notch | If Notch filter was used on the original signal or not |
| RecordingData | Date of the signal recording |
| RecordingHour | Hour of the signal recording |
| SignalDuration\_s | Duration of the signal (s) |
| Comments | Any comments |
| LinkPaper | Mentions the paper where the data has been used/analyzed |
| BackupFolder | The folder where the backup of the old signal exists |
| StimFreq | Stimulation frequency (Hz) |
| StimAmp | Stimulation amplitude (mA) |
| StimPulseDuration\_us | Stimulation pulse duration (µs) |

Another format of an old file could be as below:

*"INFORMATION"*

*"Prova2.smr"*

*""*

*""*

*"SUMMARY"*

*"1" "Waveform" "untitled" " volt" 256 255.754476 1 0*

*"31" "Marker" "Keyboard"*

*"CHANNEL" "1"*

*"Waveform"*

*"No comment"*

*"untitled"*

*" volt" 256*

*"START" 0.00000 0.00391*

*-0.00565*

*-0.01022*

*-0.01175*

and many other formats. So far, we have identified 13 types of recording formats on WBB data base. On our github repository, The matlab script *FileAdaptor.m* reads the metadata file and converts the old files for these 13 cases to the new file format such as the one below:

*"Fs= " 512*

*1.928736*

*-0.182643*

*0.024379*

*1.283741*

*.*

*.*

*.*

*-0.374178*

and names it by the unique name PID0014\_SID00127 so that all the information about the signal and subject can be retrieved from the metadata file. The new file will be saved in the desired destination. The first row indicates that the sampling frequency (*Fs*) is 512 Hz, and the rest of the rows contain the signal values at each time sample. (provide figure)

## Estimate the signals gain

In some cases, the gain of the signal i.e., the value that the recording device has used to record the signal was not known. However, we know that for example dividing the value of the signal by gain for LFPs should be in microvolts (µV) order of magnitude. Also, in some cases it can be guessed by other existing values of the same subject (e.g., similar recordings).

*GainEstimator.m* reads the metadata file and wherever that the gain was not given, tries to estimate a gain based on the SignalType and writes it to the metadata sheet and save.

## Create IEEE JSON from the new files

For web applications, it is extremely handy to merge every signal with the corresponding information from the metadata file in one single file in IEEE JSON (.json) format.

The code *JsonAdoptor.m* is a script that merges these information and saves each new signal in .json format in parallel with .txt file that already exists. This code helps researchers and web developers along with bioengineers to save any signal in .txt format along with its information as a .json file using Matlab. (provide figure)

## Clean up the data set

There were some files in the data base with no given *Fs* and since this property is essential for any kind of data analysis, one might want to keep only the signals with given *Fs*. The other essential property of the signal that is required for data analysis is the SignalType, which indicates if it is LFP, EEG or any other type of biosignal.

In this regard, *FileCleaner.m* reads the metadata file, identifies those files that have at least both *Fs* and SignalType reported and save them in a new folder called Clean. The corresponding .json files will be saved in the new folder as well.

## Data Identification & Classification

We have categorized the signals on WBB based on two distinct classes.

*Class-1* hierarchy collects signals to different "Types":

*LFP-EEG-EMG (SignalType)*

*|*

*---------------------(BodySite)*

*| |*

*STN GPi*

*|*

*-----------------------------------------------*

*| | | |*

*Gender Medication Age DBS*

*(M,F) (on/off) | (on/off)*

*|*

*-----------------------------------------------------------(age range)*

*bleow30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, above90*

Our data classification based on class-1 follows the roots as described below:

*SignalType→BodySite→SubjectGender→M or F*

*SignalType→BodySite→SubjectDoB→age range*

*SignalType→BodySite→Medication→On or Off*

*SignalType→BodySite→Stimulation→On or Off*

Part 1 of the code *Signal\_identifier\_classifier.m* constructs folders based on class-1 algorithm, copies the files from the parent folder and pastes them to the child folders.

*Class-2* hierarchy collects signals to different "Patients":

*name1-name2-...\_nameN (SubjectName)*

*| | |*

*| the same as name1*

*|*

*----------- (SignalType)*

*LFP-EEG-EMG*

*|*

*---------------------------- (BodySite)*

*| |*

*STN GPi*

*|*

*----------------------------- (Channel-Medication-Stimulation)*

*| | |*

*Side Medication DBS*

*(L,R) (on/off) (on/off)*

Classifying our data based on class-2 follows the roots as described below:

*SubjectName**→SignalType→BodySite→Channel→R/L SubjectName→SignalType→BodySite→Medication→* *On/Off*

*SubjectName→SignalType→BodySite→Stimulation→On/Off*

Part 2 of the code *Signal\_identifier\_classifier.m* constructs folders based on class-2 algorithm, copies the files from the parent folder and pastes to the child folder. The code can be edited easily based on one's need and is written such that any other classification of interest could be added.

# Results

561 signals from 57 PD patients were identified and classified based on personal information and clinical status such as on/off medication and stimulation. This is equivalent to less than 30 hours of recordings from different patients including their LFP, EEG and EMG biosignals. The old .txt files of signals were modified and their IEEE JSON format were created. Additionally, signal gains were estimated if missing.

## Equations

Number equations consecutively with equation numbers in parentheses flush with the right margin, as in (1). First use the equation editor to create the equation. Then select the “Equation” markup style. Press the tab key and write the equation number in parentheses. To make your equations more compact, you may use the solidus ( / ), the exp function, or appropriate exponents. Use parentheses to avoid ambiguities in denominators. Punctuate equations when they are part of a sentence, as in

 (1)

Be sure that the symbols in your equation have been defined before the equation appears or immediately following. Italicize symbols (*T* might refer to temperature, but T is the unit tesla). Refer to “(1),” not “Eq. (1)” or “equation (1),” except at the beginning of a sentence: “Equation (1) is ... .”

A general IEEE styleguide is available at http://www.ieee.org/web/publications/authors/transjnl/index.html



Fig. 1. Magnetization as a function of applied field. Note that “Fig.” is abbreviated. There is a period after the figure number, followed by two spaces. It is good practice to explain the significance of the figure in the caption.

# Discussion

## Sizing of Graphics

Most charts, graphs, and tables are one column wide (3.5 inches / 88 millimeters / 21 picas) or page wide (7.16 inches / 181 millimeters / 43 picas). The maximum depth a graphic can be is 8.5 inches (216 millimeters / 54 picas). When choosing the depth of a graphic, please allow space for a caption. Figures can be sized between column and page widths if the author chooses, however it is recommended that figures are not sized less than column width unless when necessary.

## Using Labels Within Figures

### Figure Axis labels

Figure axis labels are often a source of confusion. Use words rather than symbols. As an example, write the quantity “Magnetization,” or “Magnetization *M*,” not just “*M*.” Put units in parentheses. Do not label axes only with units. As in Fig. 1, for example, write “Magnetization (A/m)” or “Magnetization (Am−1),” not just “A/m.” Do not label axes with a ratio of quantities and units. For example, write “Temperature (K),” not “Temperature/K.”

Figure labels should be legible, approximately 8 to 10 point type.

### Subfigure Labels in Multipart Figures and Tables

Multipart figures should be combined and labeled before final submission. Labels should appear centered below each subfigure in 8 point Times New Roman font in the format of (a) (b) (c).

## Referencing a Figure or Table Within Your Paper

When referencing your figures and tables within your paper, use the abbreviation “Fig.” even at the beginning of a sentence. Do not abbreviate “Table.” Tables should be numbered with Roman Numerals.

## Checking Your Figures: The IEEE Graphics Checker

The IEEE Graphics Checker Tool enables authors to pre-screen their graphics for compliance with IEEE Transactions and Journals standards before submission. The online tool is located at <http://graphicsqc.ieee.org/>.

For more information on using the Graphics Checker Tool   
or any other graphics related topic, contact the IEEE Graphics Help Desk by e-mail at [graphics@ieee.org](mailto:graphics@ieee.org).

# Conclusion

## A conclusion section is required. Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the major findings and significance of the work or suggest applications and extensions.

Supplementary Materials

All the Matlab scripts developed for this study can be found

at: <https://github.com/aafarokh/WBB_NWK>.

Acknowledgment

Write “F. A. Author thanks ... .” In most cases, sponsor and financial support acknowledgments are placed in the unnumbered footnote on the first page, not here.

References and Footnotes

## References

See the end of this document for formats and examples of common references. For a complete discussion of references and their formats, see “The IEEE Style Manual,” available as a PDF link off the *Author Digital Toolbox* main page.

## Footnotes

Number footnotes separately in superscripts (Insert | Footnote).[[2]](#footnote-3) Place the actual footnote at the bottom of the column in which it is cited; do not put footnotes in the reference list (endnotes). Use letters for table footnotes (see Table I).

# Submitting Your Paper for Review

## Using Word 6.0 or Higher

All submissions must be uploaded via ScholarOne Manuscripts. Include full mailing addresses and e-mail addresses of all authors when uploading a manuscript. In addition, designate one author as the “corresponding author.” This is the author to whom proofs of the paper will be sent. Proofs are sent to the corresponding author only.

## Review Stage Using ScholarOne® Manuscripts

Contributions to IEEE OJEMB must be submitted electronically on IEEE’s on-line manuscript submission and peer-review system, ScholarOne® Manuscripts at http://mc.manuscriptcentral.com/embs-ieee. First check if you have an existing account. If there is none, please create a new account. After logging in, go to your Author Center and click “Submit First Draft of a New Manuscript.”

There are various steps to the submission process; you must complete all steps for a complete submission. At the end of each step you must click “Save and Continue”; just uploading the paper is not sufficient. After the last step, you should see a confirmation that the submission is complete. You should also receive an e-mail confirmation. For inquiries regarding the submission of your paper on ScholarOne Manuscripts, please contact oprs-support@ieee.org or call +1 732 465 5861.

## Final Stage Using ScholarOne Manuscripts

Upon acceptance, you will receive an email with specific instructions regarding the submission of your final files. To avoid any delays in publication, please be sure to follow these instructions. Final submissions should include source files of your accepted manuscript, high quality graphic files, and a formatted pdf file. If you have any questions regarding the final submission process, please contact the administrative contact for the journal.

# Editorial Policy

Do not publish “preliminary” data or results. The submitting author is responsible for obtaining agreement of all coauthors and any consent required from sponsors before submitting a paper. IEEE OJEMB strongly discourages courtesy authorship. It is the obligation of the authors to cite relevant prior work. IEEE OJEMB expects the main source of references be archival journal articles or books or thesis. Conference proceeding papers should be minimized since many are not undergoing rigorous peer reviews.

At least two reviews are required for every paper submitted. Indecipherable English is a valid reason for rejection. There is a service available that will help you improve your English for a fee, and the link to that service can be found at <http://www.ieee.org/web/publications/authors/transjnl/index.html>.

# Publication Principles

The two types of contents of that are published are 1) peer-reviewed and 2) archival. IEEE OJEMB publishes scholarly articles of archival value as well as tutorial expositions and critical reviews of classical subjects and topics of current interest.

Authors should consider the following points:

1. Technical papers submitted for publication must advance the state of knowledge and must cite relevant prior work.
2. The length of a submitted paper should be commensurate with the importance, or appropriate to the complexity, of the work.
3. Authors must convince both peer reviewers and the editors of the scientific and technical merit of a paper; the standards of proof are higher when extraordinary or unexpected results are reported.
4. Because replication is required for scientific progress, papers submitted for publication must provide sufficient information to allow readers to perform similar experiments or calculations and use the reported results. Although not everything need be disclosed, a paper must contain new, useable, and fully described information. For example, a specimen’s chemical composition need not be reported if the main purpose of a paper is to introduce a new measurement technique. Authors should expect to be challenged by reviewers if the results are not supported by adequate data and critical details.
5. Papers that describe ongoing work or announce the latest technical achievement, which are suitable for presentation at a professional conference, may not be appropriate for publication.

References

*Basic format for books:*

1. J. K. Author, “Title of chapter in the book,” in *Title of His Published Book, x*th ed. City of Publisher, Country if not USA: Abbrev. of Publisher, year, ch. *x*, sec. *x*, pp. *xxx–xxx.*

*Examples:*

1. G. O. Young, “Synthetic structure of industrial plastics,” in *Plastics,* 2nd ed., vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.
2. W.-K. Chen, *Linear Networks and Systems.* Belmont, CA: Wadsworth, 1993, pp. 123–135.

*Basic format for periodicals:*

1. J. K. Author, “Name of paper,” *Abbrev. Title of Periodical*, vol. *x,* no. *x,* pp*. xxx-xxx,* Abbrev. Month, year.

*Examples:*

1. J. U. Duncombe, “Infrared navigation—Part I: An assessment   
   of feasibility,” *IEEE Trans. Electron Devices*, vol. ED-11, no. 1, pp. 34–39, Jan. 1959.
2. E. P. Wigner, “Theory of traveling-wave optical laser,” *Phys. Rev*.,   
   vol. 134, pp. A635–A646, Dec. 1965.
3. E. H. Miller et al., “A note on reflector arrays,” *IEEE Trans. Antennas Propagat*., to be published.

*Basic format for handbooks:*

1. *Name of Manual/Handbook*, *x* ed., Abbrev. Name of Co., City of Co., Abbrev. State, year, pp. *xxx-xxx.*

*Examples:*

1. *Transmission Systems for Communications*, 3rd ed., Western Electric Co., Winston-Salem, NC, 1985, pp. 44–60.
2. *Motorola Semiconductor Data Manual*, Motorola Semiconductor Products Inc., Phoenix, AZ, 1989.

*Basic format for books (when available online):*

1. Author. (year, month day). *Title.* (edition) [Type of medium]. *volume (issue).* Available: site/path/file

*Example:*

1. J. Jones. (1991, May 10). *Networks.* (2nd ed.) [Online]. Available: [http://www.atm.com](http://www.atm.com/)

*Basic format for journals (when available online):*

1. Author. (year, month). Title. *Journal.* [Type of medium]. *volume (issue),* pages. Available: site/path/file

*Example:*

1. R. J. Vidmar. (1992, Aug.). On the use of atmospheric plasmas as electromagnetic reflectors. *IEEE Trans. Plasma Sci.* [Online]. *21(3),* pp. 876–880. Available:<http://www.halcyon.com/pub/journals/21ps03-vidmar>

*Basic format for papers presented at conferences (when available online):*

1. Author. (year, month). Title. Presented at Conference title. [Type of Medium]. Available: site/path/file

*Example:*

1. PROCESS Corp., MA. Intranets: Internet technologies deployed behind the firewall for corporate productivity. Presented at   
   INET96 Annual Meeting. [Online]. Available: <http://home.process.com/Intranets/wp2.htp>

*Basic format for reports and handbooks (when available online):*

1. Author. (year, month). Title. Comp an y . C ity, State or Country. [Type of Medium]. Available: site/path/file

*Example:*

1. S. L. Tall een. (1996 , Apr . ). The In t r an et Archi -tecture: M a nagi ng i n f o rm at i on i n t h e ne w paradigm. Amdahl Corp., CA. [Online]. Available:<http://www.amdahl.com/doc/products/bsg/intra/infra/html>

*Basic format for computer programs and electronic documents (when available online):* ISO recommends that capitalization follow the accepted practice for the language or script in which the information is given.

*Example:*

1. A. Harriman. (1993, June). Compendium of genealogical software. *Humanist.* [Online]. Available e-mail: [HUMANIST@NYVM.ORG](mailto:HUMANIST@NYVM.ORG) Message: get GENEALOGY REPORT

*Basic format for patents (when available online):*

1. Name of the invention, by inventor’s name. (year, month day). *Patent Number* [Type of medium]. Available: site/path/file

*Example:*

1. Musical toothbrush with adjustable neck and mirror, by L.M.R. Brooks. (1992, May 19). *Patent D 326 189*

[Online]. Available: NEXIS Library: LEXPAT File: DESIGN

1. This paragraph of the first footnote will contain the date on which you submitted your paper for review. This publication has emanated from research conducted with the financial support of Science Foundation Ireland under Grant number *19/IFA/7419*. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

   The next few paragraphs should contain the authors’ current affiliations. Include detailed surface address and e-mail address only for the corresponding author. For example, F. A. Author is with the National Institute of Standards and Technology, Boulder, CO, USA. S. B. Author, Jr., was with Rice University, Houston, TX, USA. He is now with Colorado State University, Fort Collins, CO, USA. \*T. C. Author is with the University of Colorado, Boulder, CO, USA (correspondence e-mail: author@uoc.edu). [↑](#footnote-ref-2)
2. It is recommended that footnotes be avoided (except for the unnumbered footnote with the receipt date on the first page). Instead, try to integrate the footnote information into the text. [↑](#footnote-ref-3)