

A Solution for Problems in the Organization, Storage and Processing of Large Data Banks of Physiological Variables

F. Palominos, H. Díaz, F. Córdova, L. Cañete, C. Durán

Fredi Palominos*, **Hernan Díaz**, **Lucio Cañete**, **Claudia Durán**

Departaments: Mathematics, Biology, Industrial Technology and Industrial Engineering.

Universidad de Santiago de Chile

Avda. Bernardo O'Higgins 3363, Santiago, Chili.

(fredi.palominos, hernan.diaz, lucio.canete, claudia.duran)@usach.cl

*Corresponding author: fredi.palominos@usach.cl

Felisa Córdova

Director at Engineering School

Finis Terrae University

felisa.cordova@gmail.com

Abstract: The proliferation and popularization of new instruments for measuring different types of electrophysiological variables have generated the need to store huge volumes of information, corresponding to the records obtained by applying this instruments on experimental subjects. Together with this must be added the data derived from the analysis and purification processes. Moreover, several stages involved in the processing of data is associated with one or more specific methods related to the area of research and to the treatment at which the base information (RAW) is subjected. As a result of this and with the passage of time, various problems occur, which are the most obvious consequence of that data and metadata derived from the treatment processes and analysis and can end up accumulating and requiring more storage space than the base data. In addition, the enormous amount of information, as it increases over time, can lead to the loss of the link between the processed data, the methods of treatment used, and the analysis performed so that eventually all becomes simply a huge repository of biometric data, devoid of meaning and sense. This paper presents an approach founded on a data model that can adequately handle different types of chronologies of physiological and emotional information, ensuring confidentiality of information according to the experimental protocols and relevant ethical requirements, linking the information with the methods of treatment used and the technical and scientific documents derived from the analysis. Consequently, the need to generate specific data model is justified by the fact that the tools currently associated with the storage of large volumes of information are not able to take care of the semantic elements that make up the metadata and information relating to the analysis of base records of physiological information. This work is an extension of our paper [25].^a

Keywords: data models; big data; EEG data organization; physiological information, metadata.

^aReprinted (partial) and extended, with permission based on License Number 3947080516854 [2016] ©IEEE, from "Computers Communications and Control (ICCCC), 2016 6th International Conference on".

1 Introduction

This paper is an extension of [25] which delves into the specification of the different types of relevant information on the EEG records, the mathematical formalization of data and interactions expands, and indicators to quantify and predict storage requirements are proposed.

Understanding the mechanisms of human reasoning and brain functioning is a central topic in neuroscience [16] [17] [18] [30] . Technological development has multiplied alternatives of

technological tools to record brain activity. In turn, the development of computers has created new opportunities for processing and analyzing the data through different tools [19] [20] [21]. Moreover, the greater availability and lower prices of technology allow researchers to access many instrumental that was previously reserved for medical specialists, allowing researchers also raised new objectives and scientific concerns.

As a result, in addition to the generation of new knowledge, there has been a huge growth in the quantity of EEG records, generating huge volumes of research and clinical data, in centers worldwide. The enormous size of these records, evidence that it is a collection of large volumes of data in the field of big data, which to be registered in computer systems, requires enormous storage media, which escape the most feature traditional computer systems.

The origin of the electroencephalogram, an instrument that allows the capture of signals known as EEG, date back to 1875 when Richard Caton first detected electrical activity on the brain surface of animals [22]. However, the expanding use of this technology beyond the traditional medical field occurs only in the last decades together with new quantitative approaches (qEEG) allowing a deeper data comprehension beyond qualitative characterizations. [23].

The potential to use this technology spans multiple areas, such as neuromarketing [24], education [26], psychology [27], labor relations [28] and work [29] .

1.1 The EEG Data and theirs specificities

Electroencephalographic data (EEG) are the record obtained by measuring brain activity, using the instrument called electroencephalographs. The measurements are obtained through a set of electrodes that are located at certain points of the human scalp, generally non-invasively, such that each electrode receives an analog signal which is then digitized by a computer. Because of the nature of the EEG data recording and digitization procedure, the temporal resolution of the signal will only be limited by the sample rate used in the experiment.

While a good amount of knowledge has been gathered from an ample range of EEG frequency span (mainly between 0,5 and 30 Hz), which mean between the EEG wave ranges of delta and beta, the gamma band (>30 Hz) has been scarcely explored.

To have a look at gamma frequencies it is necessary to increase the sample rate up to 256 Hz to have a confident resolution measuring brain phenomena that occur at 128 Hz maximum. If we want to go further, we only need to double the sample rate to 512 Hz to have a new maximum resolution of 256 Hz for the brain phenomena.

Starting with a standard EEG configuration of sample rate 128 Hz, we can explore the initial part of the gamma oscillation between 30 Hz (the end of beta) until 64 Hz (the maximum resolution allowed by Nyquist theorem). By increasing our sample rate to 256 Hz we increased data file from 64 data points in a second to 128 data points.

By setting the sample rate to 512 Hz, we ended with 256 data points per second. The fractal nature of the EEG signal makes it only finite when our technology to study it fails to go beyond.

To reach real time performances in brain-machine communication or to study very precise stimulus-response experiments, sample data must be recorded at thousands of Hz with the purpose of seeing deeper into the different time-dimensions of the processes happening in the brain.

The application of ECG in humans, although it is conditioned to the experimental protocols of each investigation, which has management, storage and processing of data concerned, have common characteristics. Sequences series of values recorded during application of an electroencephalogram (EEG), can be described as a set of analog sequences, called channels, such that each channel corresponds to record one of the electrodes of the electroencephalograph. The number of channels varies according to the instrument. The duration of each record depends on the experimental design and conditions. However, each channel will have the same duration as the

remaining channels of the RAW register.

The storage format may vary depending on the instrument type. In the case of the data corresponding to this job, they are stored in EDF format [1]. The length of each record should be as the experiment or clinical study established.

Due to the nature of brain activity, and the scope of the EEG technology, RAW records are not completely clean and can be affected by different perturbations. In the clinical field are called artifacts (residual noise from the instruments, interactions between channels, and twitching or muscle movements).

1.2 Problems on the EEG Data Management

Regarding the different factors related to storage, organization, processing and analysis of EEG records, it is clear that this is a complex problem of data management. The constant and often excessive amount of growth data and great accumulation of metadata generated in the processing and analysis of raw records. Loss of the meaning of information as a result of growth in the amount of data that over time hidden links between data, data processing, and results.

Many physiological data should be organized according to the time in which they originate, whether as instrumental records or as a result of processing. Therefore, proper management of the timestamps associated with the data is required.

In order not to incur losses of information and meaning, it is required to manage data in conjunction with information on related research, researchers, experimental people, derived data and results. Moreover, it is necessary to ensure the confidentiality of information.

Finally, it is necessary to record information about the nature of the data, their growth potential, the way they have been processed and the results of the analyses performed.

1.3 The EEG records and their relation to the area of Big Data

Current alternatives to storing and distributing data from the scope of Big Data are still weak in their ability to include metadata and meaning.

'Big Data refers to enormous amounts of unstructured data produced by high-performance applications falling in a wide and heterogeneous family of application scenarios' [7]. On the other hand, the problem of big data can be discerned as two distinct problems: Big data collections and Big Data objects [8]. 'There are three fundamental issue areas that need to be addressed in dealing with big data: storage issues, management issues, and processing issues' [12] and Big Data features can be summarized as follows: Data volume; Data velocity; Data variety; Data Value and Complexity [12] [10].

The problems in the area of Big Data arise mainly from the needs of scalability, the massive scale of the data, the heterogeneity of information, unstructured data and their distribution across multiple platforms. Resulting in problems of management, storage, portability and processing of information. Because data can be distributed across multiple sites, there are incompatibilities in the interfaces, in the definition and representation of data, which are often connected to the platform containing them. Furthermore, there are many metadata respect of the information and problems arise in the transmission of data over networks [8]. Moreover, "Big Data Also Brings About New Opportunities for discovering new values, Helps us to gain an in-depth understanding of the hidden values and incurs also new challenges, for example how to effectively organize and manage such datasets" [9].

Advances in information technology (IT), the rapid growth of so-called cloud computing and the Internet of Things (IoT) [9] have increased opportunities to generate and accumulate data, exceeding the capacity of researchers and technology companies to respond to this problem with a systematic and integrated solution. Until 2003, "five exabytes (10¹⁸ bytes) of data

were created by human. Today this amount of information is created in two days. Big Data requires a revolutionary step forward from traditional data analysis, characterized by its three main components: variety, velocity, and volume " [10] . "For solutions of permanent storage and management of large-scale datasets disordered, distributed file systems [24] and NoSQL [26] databases are good choices" [9].

McKinsey Global Institute [11] , the potential of Big Data is mainly in five sectors: Healthcare, Public Sector, Retail, Manufacturing and Personal Location Data. Moreover, the research trends in the field of Big Data analytics, point to the heterogeneity of the data and subsequent incongruity that occurs in highly unstructured data; the scalability; the combination of RDBMS and NoSQL Database Query Optimization Systems Issues in HiveQ [7].

Experience with EEG records and derived information concerning this work shows that we are in the presence of a problem of massive data rather than a problem of Big Data, however, they share many of the characteristics and problems associated with big data. In fact, the volume of data not only increases but that each action generates new records or associated metadata that systematically increase storage needs. Moreover, when there is greater availability of instrumental EEG, the growth rate of the volume of data increases linearly in proportion to the increased availability of devices.

Treatment, storage organization and EEG records and their analysis, share the problems inherent of Big Data, in several aspects:

- There are differences between different instruments to capture EEG records and have a high degree of inconsistency.
- The EEG data and its derivatives within the scope of those who is called Big Data Collections [8]. EEG data management is fully compatible with the integration of RDBMS, particularly in the management of metadata to facilitate analytical EEG records.
- EEG records management and its derivatives is a data management problem high scalability. In many types of research, data usually are not discarded but remain stored for future reference, links or reviews.

Although there storage formats for EEG records, such as the European Data Format (EDF), these records are rather unstructured nature. "In order to design meaningful analytics, it is mandatory that big data input sources are transformed into a suitable, structured format" [7].

2 Stages in the treatment of EEG records

The processing and analysis of EEG data go through different stages or states. Initially, after his capture, the EEG records untreated (called RAW) can contain different types of disturbances (artifacts), which should be identified and subtracted from the register, before applying methods of analysis. Then the treatment of previously released data, which will be subjected to different methods, create new records of information (in digital format), which are the input stage of analysis and preparation of scientific reports and clinical report. The different stages being experienced treatment EEG records can be seen in Figure 1.

The following sections provide a brief description of each of the stages or states of the processing and analysis of EEG registrations.

2.1 Data cleaning

The aim the cleaning process is the removal of artifacts [2]. The results of the data cleaning process have important effects on EEG records both in its duration as its size. During the

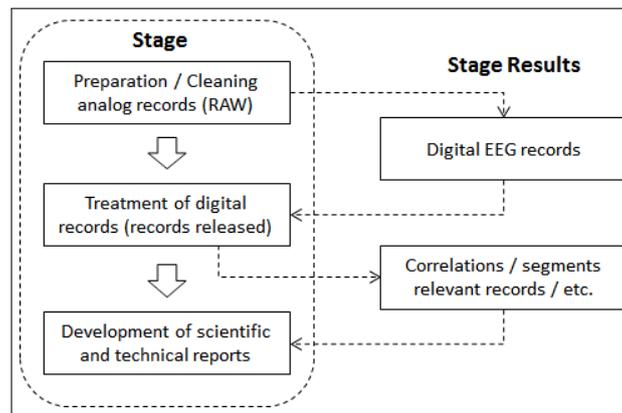


Figure 1: Steges in the treatment of EEG records

process, the analog sequences captured in each of the channels (RAW) are converted to digital streams. Thus, the accuracy and relevance of the digital signal generated will depend on the method parameters that are applied in the purification process.

There are different types of alternative processes for the cleaning of raw records. Each EEG record may be subjected to a sequence of various cleaning methods. Some of these results will be discarded during treatment and others will be stored for later analysis. In order to properly interpret the results of the clearance records, you must record the sequence of the methods applied, and the specific values of the parameters that each of the methods was applied.

In the case of EEG records considered in this work, methods and purification processes used are the following:

- **Importing Data:** Due to the Emotiv EPOC instrument records the sensor signals in 14 useful channels (called: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4), a file format of European data format (EDF), the import is done through the built-in tool BIOSIG EEGLAB [4] software. Generally, you must import all channels, but there is the possibility of importing a few.
- **Epoch's selection:** It refers to select periods of records that contain events of interest to the investigation, in which people objects of study, they are subject to some kind of stimulus.

At this stage, you can also perform other actions considered pre-processing such as changing the sampling rate, filter the data or re-reference them.

2.2 Processing and data exploration

It is understood by data processing, to all actions that aim to inquire into the information contained in the records or derive new information from existing data. There are different methods and techniques that can be used at this stage [4]. The specific order in which they are used, the number of times that it was applied as well as its parameters, depend on the objectives of the research and the findings that researchers performed during the same.

Processing methods according to their purpose and their effect on the data can be classified into:

- **Exploratory methods:** Specifically, rather than processing methods they are techniques to explore the characteristics and peculiarities of each EEG record. Generally make use of the help of graphical tools that allow the display of one or more channels, thus forming a

better impression and understanding of the records. Including loading and channel display (load and view channel locations), ERP Plotting images, and Plotting Spectra and Maps Channel [4] are frequently mentioned.

- Methods of processing and generation of macro data: It refers to those methods that act on the data to modify or to generate new data. Within this group are, for example, those for determining the baseline (to avoid skewed by the presence of low-frequency artifacts or analysis), ICA using data decomposition [2], work with ICA components, the decomposition time / frequency [4], etc.

3 Context of the capture, storage, processing, and analysis of EEG data

Scientific and clinical studies that require registration, treatment, storage and analysis of EEG data, demand the articulation of investigative processes as well as different types of elements, such as clinical patients, volunteers, premises, equipment, technical staff, clinicians, and researchers (see figure 2). The relevant part of these activities corresponds to drawing conclusions: technical, scientific and clinical diagnostics, which must be properly documented.

The intervention of so many different factors, not only generates abundant records of information of different nature but also requires an organization and storage that includes the different interactions between different components, in order to avoid levels of confusion and distancing from reality, a product the abundant data.

This paper addresses the definition of a data model (see figure 4) that supports the development of an information system based on a storage strategy and organization that considers existing relevant interactions in the context of studies related to the EEG data.

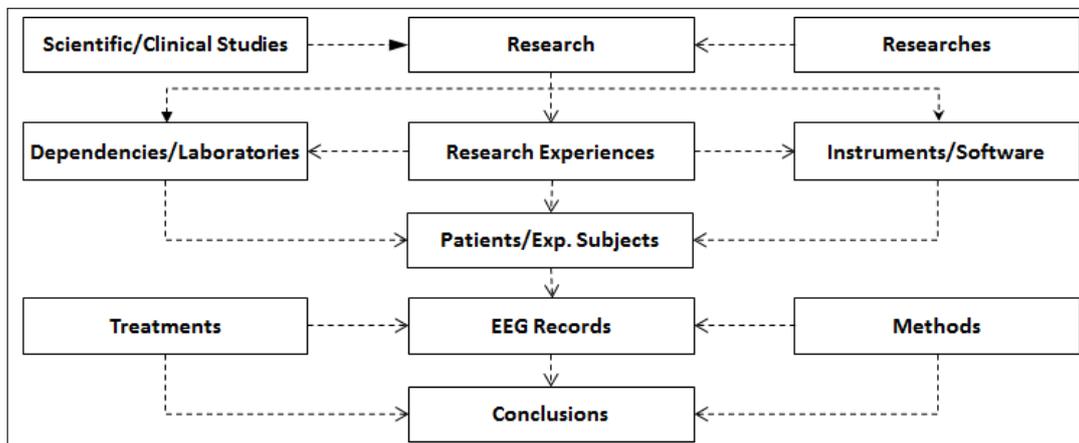


Figure 2: The context in which they exist and the EEG data is generated

The following sections describe the conceptual data model developed to meet the requirements described previously.

3.1 Scientific/clinical studies, research and dependences

In any context of treatment and analysis of EEG data, there will be planned investigative processes (projects) conducted by researchers or clinical staff specialized, which will be assigned to a set of physical units (clinics or laboratories), where all the experimental activities will be

conducted. The enclosures also house the instruments and computers with the necessary software for the treatment of captured EEG records.

3.2 Research experiences and records

The projects require the development of various clinical or experimental activities, which are subject to strict protocols, which therefore must be developed, conducted and supervised by qualified researchers and clinicians. The result of these activities is the generation of several EEG records (RAW), which will be generated in each of the sessions which will be subjected patients or experimental subjects. Therefore, it comes to many EEG records, which require adequately linked to experiences in which they originate, the instruments used, laboratory personnel and dependencies in which they perform.

3.3 Dimensioning and predicting storage needs

The storage, management, processing, and documentation of the results of processing and analysis of EEG records obtained from research and clinical studies, requires a hardware platform that, in addition to processing power, ensure responses in appropriate time lapses and adequate storage management. Investigations that require storage of EEG records and derivatives of them require computers able to process and manage files that may have average sizes of hundreds of megabytes.

The type of study that originates the data, affect considerably on the number of records, duration, and size. In the medical field, the amount, duration, and size of EEG records is directly related to the pathology under study and its severity. Experts know beforehand the size and approximate duration of EEG records that are necessary to diagnose with a significant degree of reliability of the presence or absence of certain pathologies.

However, in the field of new applications of EEG technology, which have opened to very different areas outside of classical medicine, the problems have storage requirements and processing needs very variability. Any useful prediction requires the determination of significant variables that serve as the basis for accurate estimates.

Suppose that S is the set of study areas related to EEG data and is I_S the set of research (projects) linked around these areas, then let:

- $P(I)$: The set of all persons involved in a research I , $I \in I_S$.
- $R(x, I)$: The set of EEG records captured a person x in an investigation I , $I \in I_S$ and $x \in P(I)$.
- $RD(x, t, r)$: The set of records derived from treatment of a record EEG r obtained from a person x in research I , $I \in I_S$, $x \in P(I)$, and $r \in R(x, I)$.

If we denote by $I(r)$ to all of the records included in a full investigation I , then the storage space of research related to the different areas of study contained in S , is given by the following expression:

$$Z(S) = \sum_{I \in I_S} \sum_{x \in P(I)} \sum_{r \in R(x, I)} (Z(r) + \sum_{r_d \in R_d(x, I, r)} Z(r_d)) \quad (1)$$

where $Z(x)$ represents the size of an object x measured in MB.

To study this problem, It has been chosen a set of indicators whose behavior is expected to permit characterize the different types of studies and estimating storage needs requiring treatment and analysis of neurophysiological information.

The indicators initially considered are as follows:

Indicator	Description
MB	Amount of storage required by RAW records and derivatives in each investigation.
NS	Amount of subjects involve in each investigation.
NR	Amount of records RAW captured in each investigation.
ND	Amount of (digital) derived records RAW records in each investigation.
MS	The total sum of the minute records captured in a research.
NI	Amount of reports generated in each investigation.

Table 1: Indicators relating to a representative set of EEG investigations.

3.4 Organization of processing methods and reports

One of the ultimate goals of the research is drawing conclusions and scientific / technical reports. This requires the systematic application of various methods of processing and analyzing the results, which will provide researchers with lights on the nature of the phenomena under study.

The effectiveness and relevance of each method are directly related to its parameters and the order or sequence in which they are applied. In many cases, each method may require multiple applications to find a suitable parameterization for the problem, something that must be recorded in the study as a log for further review.

The characterization of a processing method m any can be represented by $m(n, \vec{V})$, where n represents the number of parameters involved in the application of the method and \vec{V} is a vector containing the formal description of parameters.

Processing methods are applied to the original EEG records (RAW) or records arising from the application of previous methods (eg cleaning artifacts). The fact of applying a processing method on a specific EEG record r , can be represented by the cinchona (m, r, t, n, \vec{v}) , such that the vector \vec{v} contains the specific values of the parameters applied to register r at time t , which can give rise to a new data record r_t .

Thus, the full process a record EEG r can be represented as the set

$$(m(n_i, \vec{V})_i, r_i, t_i, n_i, \vec{v}_i) / i = 1 \dots k \quad (2)$$

We denote as $M(r)$ a sequence of methods applied in a temporal sequence t_1, \dots, t_k over an r record.

Given the research process, I involves the generation of one or more records for each individual, we represent by $M_x(r)$ the set containing all sequences $M(r)$ associated with all EEG records applied to the subject S .

Finally, the results obtained from the analysis of all processing sequences contained in $M(r)$ in a research I , give rise to many technical documents or scientific reports that should be properly cataloged and stored for any use present or future.

Scientific and clinical reports contains the conclusions of the analysis made in the investigation and therefore should be linked to each of the records and data used in its construction.

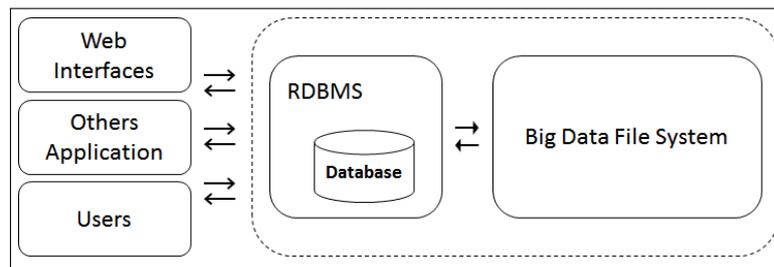


Figure 3: Interaction between system components

4 Data model for the organization, storage and processing of large data sets EEG

It needs to consider the context in which EEG records are recorded, processed and analyzed from multiple investigations; consider that the data from the processing of each of the records are in the presence of a problem of management of complex data, which also requires the coordination and integration of multiple sources of information and technologies. To keep the whole meaning of the information is necessary to include descriptive information concerning research, technical reports, and the conclusions drawn from the records and the various associations that occur between these entities.

The EEGLAB [3] [4] tool has addressed this problem from the standpoint of ease of processing and exploration of the data but does not provide for integration into the information system of the context in which the records are generated. Neither includes information on technical documents with the conclusions drawn from his treatment.

This paper presents a conceptual model for the organization, storage and processing of EEG records (under the Model Entity / Relationship [5]), which takes over the metadata associated with the problem. Considers context information where records are generated. It takes over the need to store the associations between records, treatment, scientific papers and technical reports, which contain the findings of the analysis developed.

This proposal consists in adding information on the context in which data are generated to information concerning the organization, storage and processing of EEG records. This is achieved by a technology-based relational database component, which incorporates contextual information and semantic elements that give meaning to data volumes (see figure 4). Subsequently, on this platform may implement other applications, such as web-based interfaces or other tools that use the metadata system for maintaining and managing the links between different elements of information required to manage information EEG.

4.1 Scientific/clinical studies, research and dependences

The data EEG focus of this study is captured in scientific or clinical studies during the development of properly planned processes (projects), conducted by researchers or clinical professional, in installations or laboratories that house the instruments required to capture data and experimentation. The projects are framed within these studies, and in turn, researchers or clinicians may be linked to different projects. The diagram is shown in Figure 4 depicts this dynamic. In particular, it must be considered clearly states that all projects should be assigned to a field concerning clinical trials or scientific experience, which comprise a wide range of investigations.

4.2 Research, experimental situations, experimental subjects and EEG records

Normally, an investigation involves making different experimental situations or activities, which can obtain new EEG records of experimental subjects, using various instruments. Then each EEG record, although will be associated with only one subject and a specific instrument can be used in other investigations. A similar situation applies to clinical records obtained from patients.

This situation creates a set of EEG records, previously termed $I(r)$, which can be represented as in Figure 4 by record entity.

In research contexts addressed in this work, records obtained from the application of an individual experience, are always linked to a single instrument. However, you can simultaneously record multiple records across different instruments. This situation also has been reflected in Figure 4 as an aggregation (high-level entity) between Experience and Instruments entities.

On the other hand, as shown in Figure 4, the association between a Research and high-level entity Experience/Instruments, is the type many to many, because an investigation can involve more than one pair Experience/Instruments, and in turn, each pair experience/instrument may be associated with multiple investigations. This relationship between Experience/Instrument and Research frequently applied to clinical studies where the same experimental situation, for example, diagnostic procedures are applied to multiple patients. On the other hand, people involved in an investigation or are subjected to a clinical procedure may eventually be required to participate in other research or studies.

4.3 Methods, sequences processing, and parameter settings.

The systematic and successive application of various treatment methods on EEG records in RAW format or those derived from previous treatments is an essential feature of EEG data analysis. Though several treatments do not give useful results, many treatments if they are significant for research, so that should be recorded, including parametrization, identifying the set of input data, and the order and the moment in time they were applied. Consequently, in the exploration and processing of each EEG recording, you must register the application sequence of processing methods. The explorations on the results, all associated with a set of timestamps. Processing said the sequence of previously denoted by the expression (2).

These timestamps introduce a natural chronological order that besides being a historical record of the treatment records, allows us to analyze the strategy analysis and correct it, if necessary. This situation it's represented in Figure 4. The application of different parametrized methods results in the generation of new data records, which add to existing records. This situation is collected as a reflexive relationship called "originates", between the high-level entity called "applies" and the entity called Record.

4.4 Management of technical reports

It is considered that, in general, an investigation involves different experimental situations, which are applied to different people. This creates a set of EEG records, previously has been called $I(r)$.

The processing of EEG records, either through exploratory methods or data processing allows obtaining derived information that is the basis of the conclusions contained in the technical and scientific reports. The preparation of the documents can take information from many records corresponding to a specific research process or multiple investigations, a reason that is it necessary to link the EEG records all the scientific/technical reports involved.

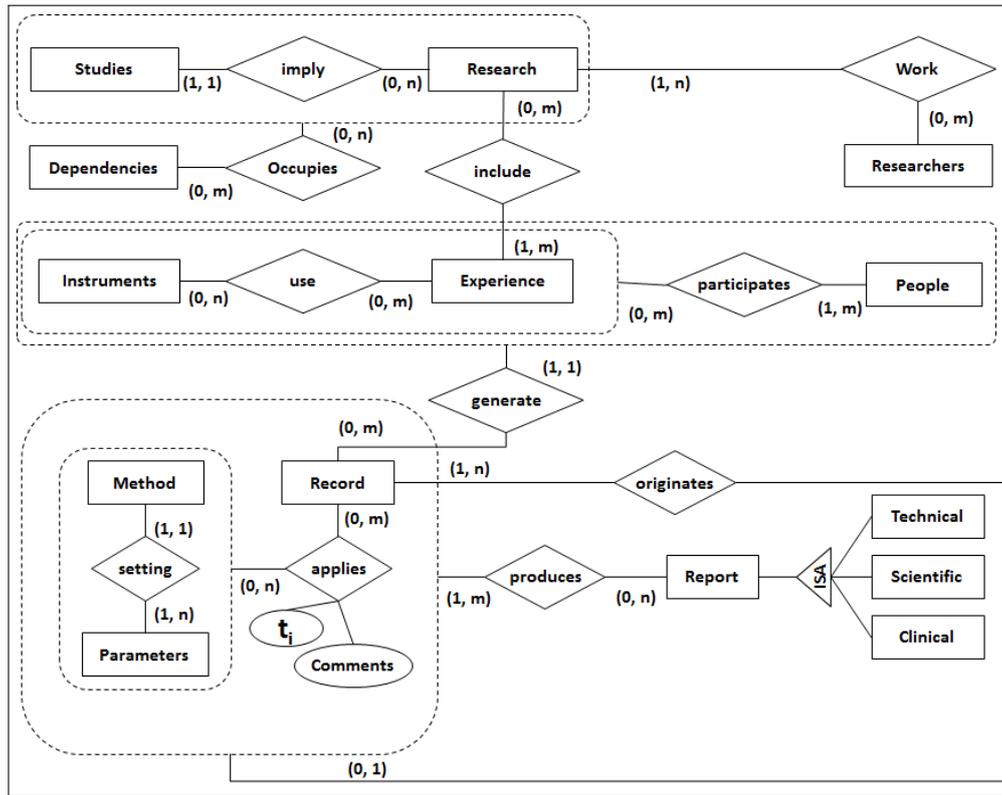


Figure 4: Conceptual model for the context of EEG data

The biggest challenge in the treatment of biometric information is not only the need for storage space for RAW records or derivatives but also store reports and other results containing the summary and details of the conclusions.

4.5 Overview of the model for the management, storage and processing of EEG records

Integrating all requirements described in the preceding paragraphs, is represented in the conceptual data model is presented in figure 4. To simplify the figure, the attributes have not been incorporated into the model.

This data model aims to implement a database (with documentary features), to incorporate contextual information and references to the physical location where the data is stored (URL). Specifically, the purpose of the database is as follows: (1) Integrate records and metadata to facilitate the analytic EEG records and derived data; (2) Establish a catalogue of EEG records, scientific and technical documents as well as from the investigation; (3) Allow the generation of a system capable of recording the activities of research centres and evolve if necessary; and (4) Allow the development of tools for analysis and processing of data, which are made in the context of the system, automatically integrating partial and final results of interest to research centres.

To facilitate the implementation of the de model in any relational platform and therefore the portability of applications that make use of it, in figure 5 shows the standard logical design (DLS) of the conceptual model.

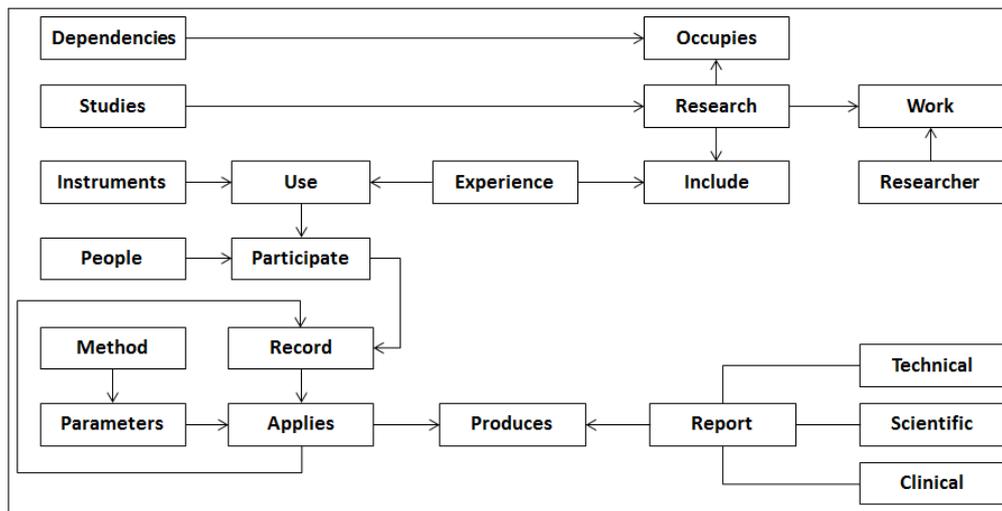


Figure 5: DLS for the conceptual model about context of EEG data

Conclusions

The growing needs to store large amounts of information relating to various areas of human and scientific work, has prompted different approaches and approaches related to the storage of information, is now called Big Data. Although these proposals address many of the problems associated with the storage, access, and distribution of information on computer media, they do not have sufficient tools to manage the semantic elements of the stored information, which is specific to each problem type. In the case of EEG records and technical and scientific reports, it is required of mechanisms for the organization, storage and processing of data, allowing the integration of the various elements of information contained in the records and mainly those generated from their treatment.

Current developments in the area of storage [13] [14] provide an alternative for managing large volumes distributed in different storage media, which however are limited by the heterogeneity of the different types of data. Proposals made in this document provide a solution for these semantic constraints in the context of specific data, by integrating other technologies, such as database services (DaaS) [15] [7], which also allow manipulation of metadata and integrate different types of information, enabling the development of applications and interfaces that automate the recording and metadata generation. In addition, the technology database allows inheriting mechanisms to ensure the confidentiality of information regarding experimental subjects, the analysis, and conclusions of the records, research, and researchers.

In this scenario the data model proposed is able to leverage solutions in the field of Big Data, incorporating semantic elements that enable researchers and technical staff not to lose control over data and new knowledge derived from them. The data model becomes a significant interface between the information processing methods, facilitating data analysis and generation of conclusions. Moreover, the increased availability of metadata and recording timestamps on different moments related to the processing, analyzing and drawing conclusions, promote more efficient use of computing resources by decreasing the need to reprocess data unnecessarily when those results are registered for the system level.

Finally, the use of technologies of conventional databases available in the field of free software or proprietary contributes to the greater integration of information value in the new scenario underlying the current developments in the field of human-computer interfaces in science and medicine.

Acknowledgment

This work was developed under the project "Definición, desarrollo e implementación de nuevos indicadores y procedimientos para la evaluación y seguimiento del rendimiento académico estudiantil y el desempeño docente, en base información curricular, variables psico-neuro-cognitivas y herramientas basadas en matemática y lógica difusa", which receives funding from the Department of Scientific and Technological Research (DICYT) of the University of Santiago of Chile (USACH), Chile.

Bibliography

- [1] European Data Format (EDF). [En línea], Disponible at <http://www.edfplus.info/> [Accesed: 17-nov-2015].
- [2] T.-P. Jung et al. (1998); Removing electroencephalographic artifacts: comparison between ICA and PCA, *Neural Networks for Signal Processing VIII, 1998. Proceedings of the 1998 IEEE Signal Processing Society Workshop*, 63-72, DOI: 10.1109/NNSP.1998.710633
- [3] A. Delorme, S. Makeig (2004); EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis, *Journal of neuroscience methods*, 134(1): 9-21, 2004.
- [4] EEGLAB Tutorial: Table of Contents, <http://sccn.ucsd.edu/eeglab/eeglabtut.html>
- [5] Chen Peter (1976), The Entity-Relationship Model - Toward a Unified View of Data, *ACM Transactions on Database Systems*, 1(1): 9-36.
- [6] E.F. Codd (1970); A relational model of data for large shared data banks, *Communications of the ACM*, 13(6):377-387.
- [7] A. Cuzzocrea, I.-Y. Song, y K. C. Davis (2011); Analytics over large-scale multidimensional data: the big data revolution, *Proceedings of the ACM 14th international workshop on Data Warehousing and OLAP*, 101-104.
- [8] M. Cox, D. Ellsworth (1997); Managing big data for scientific visualization, *ACM Siggraph*, 97: 21-38.
- [9] M. Chen, S. Mao, Y. Liu (2014); Big Data: A Survey, *Mobile Networks and Applications*, 19(2): 171-209.
- [10] S. Sagirolglum, D. Sinanc (2013); Big data: A review, in Collaboration Technologies and Systems (CTS), *2013 International Conference on*, 42-47.
- [11] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh and A.H. Byers (2011); *Big data: The next frontier for innovation, competition, and productivity*, McKinsey Global Institute, 2011.
- [12] S. Kaisler, F. Armour, J. A. Espinosa, W. Money (2013); Big Data: Issues and Challenges Moving Forward, *2013 46th Hawaii International Conference on System Sciences*, 995-1004.
- [13] Zikopolous Paul, Deroos Dirk, Deutsch Tom, Lapis George (2012); *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*, McGraw-Hill, 2012.
- [14] HPCC Systems. [Online]. <https://hpccsystems.com/>

-
- [15] H. Hacigumus, B. Iyer, S. Mehrotra (2002); Providing database as a service, en Data Engineering, *2002 Proceedings. 18th International Conference on*, 29-38.
- [16] B.B. Biswal et al. (2010); Toward discovery science of human brain function, *Proceedings of the National Academy of Sciences*, 107(10): 4734-4739.
- [17] A. Paul Alivisatos et al. (2012); The Brain Activity Map Project and the Challenge of Functional Connectomics, *Neuron* 74(6): 970-974, DOI: <http://dx.doi.org/10.1016/j.neuron.2012.06.006>
- [18] Monti M. M., Parsons L. M., Osherson D. N. (2009): The boundaries of language and thought in deductive inference, *Proceedings of the National Academy of Sciences*, 106(30): 12554-12559.
- [19] Evan Heit (2015); Brain imaging, forward inference, and theories of reasoning, *rontiers in Human Neuroscience*, January 2015, Vol. 8, Article1056.
- [20] Vinod Goel, Brian Gold, Shitij Kapur (1997); Sylvain Houle, The seats of reason? An imaging study of deductive and inductive reasoning, *NeuroReport*, 8: 1305-1310.
- [21] Ranjit A. Thuraisingham, Georg A. Gottwald, On multiscale entropy analysis for physiological data, *Physica A: Statistical Mechanics and its Applications*, Volume 366, 1 July 2006, Pages 323-332.
- [22] Hass L.F. (2003); Hand Berger (1873-1941), Richard Caton (1842-1926), and electroencephalography, *Journal of Neurology, Neurosurgery and Psychiatry*, 74(1):9, doi:10.1136/jnnp.74.1.9.
- [23] Kececi H., Degirmenci Y. (2008); Quantitative EEG and cognitive evoked potentials in anemia, *Clinical Neurophysiology*, 38(2): 137-143, doi:10.1016/j.neucli.2008.01.004
- [24] Rami N. Khushaba et al. (2013); Consumer neuroscience: Assessing the brain response to marketing stimuli using electroencephalogram (EEG) and eye tracking, *Expert Systems with Applications*, 40:803-3812.
- [25] Fredi E. Palominos, Hernan Diaz, Felisa M. Cordova, Cañete Lucio, Claudia A. Duran, Model for the organization, storage and processing of large data banks of physiological variables, *Computers Communications and Control (ICCCC), 2016 6th International Conference on*, e-ISBN 978-1-5090-1735-5, IEEE Xplore,173-179, DOI: 10.1109/ICCCC.2016.7496757
- [26] Díaz, Hernán, L. Cañete, F. Palominos, C. Costa y F. Córdova (2012); Neurotechnologies for Education Improvement: Self-Knowledge After Opening the Black-box, *Conference Proceedings of the International Symposium Research and Education in Innovation Era. ISREIE, 4th Edition Arad. Journal Plus Education*, 8(2): 44 - 52.
- [27] Paul Howard-Jones et al. (2013); Neuroscience and Education: Issues and Opportunities, TLRP Teaching and Learning Research Program; ESRC Economic and Social Research Program, University of London, ISBN: 0-85473-741-3, 2013.
- [28] Maria Kozhevnikov (2007); Cognitive Styles in the Context of Modern Psychology: Toward an Integrated Framework of Cognitive Style, *Psychological Bulletin*, 133(3): 464-481.
- [29] Matthew D. Lieberman (2007); Social Cognitive Neuroscience: A Review of Core Processes, *Annu. Rev. Psychol.* 58:259-289.

- [30] Raja Parasuraman, George Mason, Glenn F. Wilson (2008); Putting the Brain to Work: Neuroergonomics Past, Present, and Future, *Human Factors*, 50(3): 468-474. DOI 10.1518/001872008X288349.