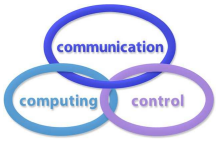
INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
Online ISSN 1841-9844, ISSN-L 1841-9836, Volume: 21, Issue: 6, Month: December, Year: 2026
Article Number: 7525, https://doi.org/10.15837/ijccc.2026.6.7525



CCC Publications



Contextualizing EEG-Derived Results: A Semantic and Multidimensional Architecture for Cross-Study Analysis

F. Cordova, F. Palominos

Felisa Margarita Cordova Gonzalez*
Full Professor
Facultad de Ingeniería
Universidad San Sebastian
7 Bellavista, Recoleta
Santiago, Chile
*Corresponding author: felisa.cordova@uss.cl

Fredi Edgardo Palominos Villavicencio
Full Professor
Departamento de Matemática y Ciencia de la Computación
Facultad de Ciencia
Universidad de Santiago de Chile
165 Las Sophoras, Estación Central
Santiago, Chile
fredi.palominos@usach.cl

Abstract

The capture, processing, and analysis of electroencephalographic (EEG) recording data from different experiments, the application of processing methods, and different levels of temporal and spatial granularity are usually stored and analyzed in isolated contexts, making it difficult to trace, compare, and reuse them. This proposal builds upon experience gained from empirical studies involving cognitive assessments, problem-solving tasks, mental imagery experiments, correlation and synchrony analyses, the Hilbert–Huang Transform, and Hurst exponent analysis. These studies reveal a common multidimensional structure that can be leveraged to organize EEG-derived information according to subjects, experiments, methods, channels, brain regions, time scales, and experimental conditions. This work proposes an integrated neuroinformatics architecture for managing, analyzing, and leveraging knowledge derived from EEG data. It integrates a repository of raw EEG data; a semantic layer linking experiments, subjects, methods, processing conditions, and derived results; and a multidimensional environment based on OLAP structures to organize and compare heterogeneous analytical results. Built upon this integrated structure, the proposal includes an AI-supported knowledge management layer designed to identify regularities among accumulated results, contextualize detected patterns using semantic information, and generate candidate knowledge for subsequent validation by specialists. Validated knowledge can be progressively accumulated and reused to interpret new results and support future analyses. The architecture preserves traceability from the generated knowledge back to the analytical results, methods, experiments, and raw EEG data. It also allows for the incorporation of various artificial intelligence

and machine learning techniques tailored to the specific characteristics of each analytical problem. The proposed framework offers an extensible foundation for the cross-study analysis of EEG data, reproducibility, and the progressive accumulation of knowledge in neuroinformatics.

Keywords: : EEG analysis; multidimensional analysis; cognitive neuroscience; data modelling; brain–computer interface; knowledge management; language models; OLAP model.

1 Introduction

Understanding how the human brain organizes, processes, and integrates information remains one of the most complex challenges in neuroscience [2], electroencephalography (EEG), as a noninvasive technique for recording cortical electrical activity, has become an essential tool for exploring neural dynamics underlying cognition, emotion, and behaviour. EEG's millisecond temporal resolution allows the detection of fast oscillatory phenomena, providing valuable insights into the synchronization and connectivity among different cortical regions [1]. In recent years, the intersection between EEG signal analysis, cognitive modelling, and computational data science has opened new possibilities for identifying functional brain patterns that correspond to learning processes, problem-solving, and imaginative activities [4], [5], [13].

The increasing availability of portable and wireless EEG systems, such as the Emotiv Epoc, has democratized access to neurophysiological data, making it feasible to integrate brain monitoring into research on education, mental health, and human–computer interaction. However, the exponential growth in data volume and complexity presents major methodological challenges [3]. Traditional EEG analysis methods—typically based on linear and stationary assumptions, such as Fourier Trans form—fail to capture the nonlinear, dynamic, and non-stationary nature of cerebral activity. To ad dress these limitations, advanced approaches such as the Hilbert–Huang Transform (HHT) and Hurst exponent analysis have emerged as powerful tools for exploring nonlinearity, long-memory processes, and order–chaos dynamics in brain signals [6],[9],[12][14].

The increasing use of instruments to measure different electrophysiological variables has highlighted the use of EEG in cognitive neuroscience and applied neurophysiology. From a cognitive perspective, learning styles and problem-solving behavior have been studied through psychological frameworks such as Kolb's experiential learning model, the VAK (Visual Auditory–kinesthetic) model, and Herrmann's brain dominance theory. Yet, these theoretical constructions have often lacked objective neurophysiological correlates [5],[14],[20]. The convergence of cognitive testing (e.g., Raven's Matrices) with EEG-based measures of brain connectivity offers a unique opportunity to bridge behavioral assessment and cerebral function [18],[20]. Exploring how individual differences in cognitive style correspond to neural synchrony patterns may improve the understanding of how knowledge is acquired, processed, and represented in the brain [26],[27].

Multidimensional data management model based on OLAP technology was designed to integrate raw EEG data, derived variables, and metadata, ensuring semantic traceability across analytic procedures. By applying Online Analytical Processing (OLAP) and multidimensional database technology, EEG data can be organized not only by temporal or spectral parameters but also by cognitive task, participant characteristics, and analytical outcomes [16],[17]. This integration allows dynamic querying, cross-study comparison, and intelligent visualization of large datasets, transforming raw brain recordings into structured, interpretable knowledge.

The integration of EEG signal analytics with neuroinformatic principles represents a crucial step towards scalable and reproducible neuroscience. Neuroinformatic seeks to combine electrophysiological, behavioral, and metadata layers into interoperable structures that support semantic traceability and multidimensional exploration [14]. The present work summarizes and unifies experimental and methodological studies conducted over several years, each contributing to the development of a comprehensive system for EEG-based cerebral data analysis. The studies range from identifying EEG correlates of cognitive learning styles to developing novel topographic visualization methods and data storage architectures. Together, they demonstrate an interdisciplinary approach combining signal processing, cognitive science, and data management engineering. The overarching aim is to construct a neuroinformatic framework that facilitates multidimensional analysis of brain activity and promotes

understanding of the neural mechanisms that support learning, imagination, and cognitive organization.

This paper therefore presents a technical–methodological approach that integrates three primary dimensions: cognitive modeling and behavioral analysis, linking psychometric assessments to electrophysiological activity; advanced EEG signal analytics, employing correlation matrices, HHT, and Hurst exponent to quantify synchrony and complexity; and intelligent data management, applying OLAP-based models to store, organize, and visualize EEG information.

Although this powerful ecosystem has allowed a better understanding of increasingly specific aspects of brain activity, a structural problem persists concerning the fact that the derived results are neither stored nor organized in an integrated manner. This hinders systematic comparison between studies, traceability of processing, joint analysis of different metrics, and analytical exploitation in both inter- and intra-subject contexts. Previous developments have primarily focused on conceptual models for RAW recordings and metadata from the capture process, leaving out the organizational structure of the complex results generated by processing. The aim of this article is twofold. First, to present a systematic review of EEG analysis methods relevant for designing storage models oriented toward analysis. Second, to propose a multidimensional (OLAP) architecture capable of integrating the most relevant derived results, allowing flexible analytical queries, temporal comparisons, aggregate analyses, and methodological traceability. The contribution focuses on providing experimental neuroscience with storage and processing contexts based on multidimensional data models, an aspect that has been explored little in the current literature.

It is a fact that the emergence of artificial intelligence opens a huge range of possibilities to significantly advance our understanding of brain activity, as well as in the interpretation of the signals produced by such activity. Our proposal falls within the study of understanding the meaning of brain activity signals. Despite the progress made, the physiological and cognitive interpretation of the observed patterns remains an open problem that requires accumulated evidence and evaluation by specialists. Research on deep learning applied to EEG shows the potential to obtain useful representations of complex signals, but it also raises demands for comparability and reproducibility. Roy et al. [22], in a systematic review of 154 studies, identify a wide diversity of data and procedures and warn that the lack of availability of data and code makes it difficult to reproduce many Results. In [24] they propose a graph convolutional network that dynamically learns the relationships between EEG channels to recognize emotions. Meanwhile, Kostas et al. in [15] present BENDR, which combines transformers and contrastive self-supervised learning to obtain raw EEG representations that can be transferred across datasets with different subjects, equipment, and tasks. Together, these works illustrate the variety of results that AI methods can produce. From this, it follows, as a requirement for the architecture proposed in this article, the convenience of linking each result with its source data, experimental conditions, processing procedures, and method used.

2 Materials and Methods

2.1 Basics and Characteristics of EEG Information

Electroencephalography (EEG) allows us to record the electrical variations generated by groups of cortical neurons. The signals obtained have a multivariate nature, strong temporal dependence, and can be influenced by both physiological conditions and environmental artifacts. Likewise, their behavior can be affected by the cognitive, perceptual, or motor tasks that the person is performing during recording with instruments. EEG signal analysis usually considers different spectral components, including the delta, theta, alpha, beta, and gamma bands as shown in Figure 1. The variations seen in these bands help characterize different conditions related, for example, to cognitive processes, attentional states, or levels of mental workload.

However, the EEG recording is only the starting point of a broader analytical process. From the original signals, different preprocessing steps need to be carried out and various analysis methods applied, which produce new sets of data. For this reason, the problem of managing EEG information cannot be limited exclusively to storing the original recordings, but must also take into account the information generated during their processing and analysis.

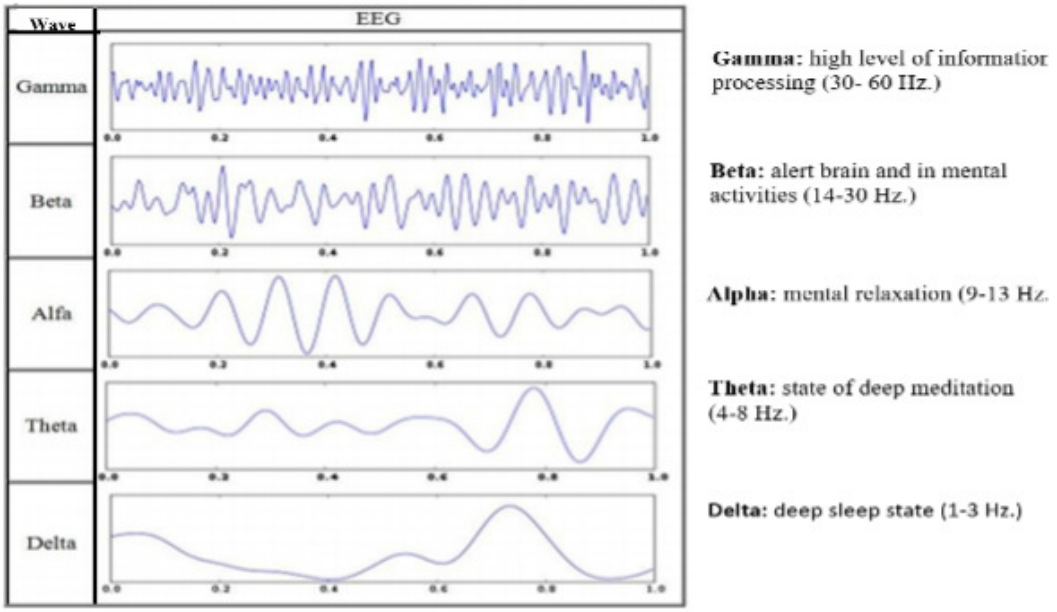
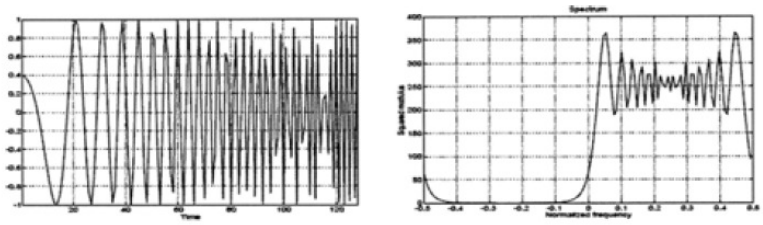


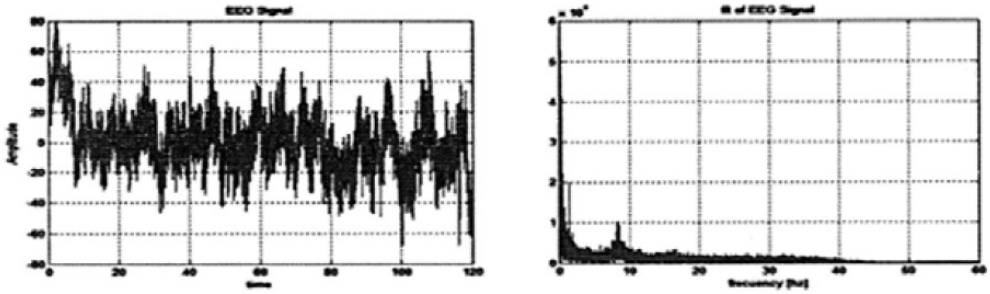
Figure 1: EEG representative waves

2.2 EEG as a multivariate time series problem

From a computational perspective, an EEG recording can be considered as a set of time series that show different degrees of correlation, various temporal and spatial scales, and nonlinear dynamic behaviors. These features require using analysis techniques capable of representing, among other aspects, temporal variations, such as moving windows or dynamic correlations; spectral variations, as well as spatial patterns, like topographies or measures of synchronization between signals. Therefore, the results obtained from EEG can show considerably different structures depending on the method used. A linear correlation, a nonlinear complexity measure, a spectral analysis, or an adaptive time-frequency decomposition produce results with different structures, dimensions, and granularities as presented in Figure 2.



(a) Linear but non-periodic signal and frequency spectrum



(b) Non-linear, non-periodic and non-stationary signal and frequency spectrum

Figure 2: Nonlinear, temporal, spectral scale and spatial patterns

2.3 Preprocessing and generation of derived information

2.3.1 EEG Acquisition and Preprocessing

Preprocessing is a fundamental step in EEG analysis. EEG data were collected using a 14-channel wireless EEG headset, with electrodes positioned according to the international 10–20 system (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4). The sampling rate ranged between 128 and 256 Hz, depending on the study configuration. The process of acquisition and preprocessing of raw EEG signals normally uses the EEGLAB toolbox in MATLAB® [8]. The preprocessing pipeline includes band-pass filtering into canonical frequency bands (δ: 0.5–4 Hz; θ: 4–8 Hz; α: 8–13 Hz; β: 13–30 Hz; γ: 30–45 Hz); baseline correction and artifact removal (eye blinks, muscle movements, electrode drift); Independent Component Analysis (ICA) for separation and cleaning of non-cerebral components; and segmentation into epochs corresponding to specific cognitive or imagery tasks.

Operations such as artifact filtering, segmentation, independent component analysis (ICA), removal of artifactual components, and conversion of recordings to standardized formats (EDF) generate additional information that is relevant for interpreting later results. The decisions made during preprocessing affect the data on which analytical methods will be applied and, therefore, are part of the context needed to interpret the results obtained.

2.3.2 Analytical Processes and Techniques

Starting from the preprocessed signals, multiple analysis techniques can be applied to obtain neurocognitive patterns and brain dominances following the system shown in figure 3. Depending on the method chosen, the results can take very different forms: correlation matrices and maps as shown in Spearman R cross-correlation map. Also, spectra, indices calculated over different scales, time series of complexity measures, component decompositions, topographic representations, or connectivity measures between channels.

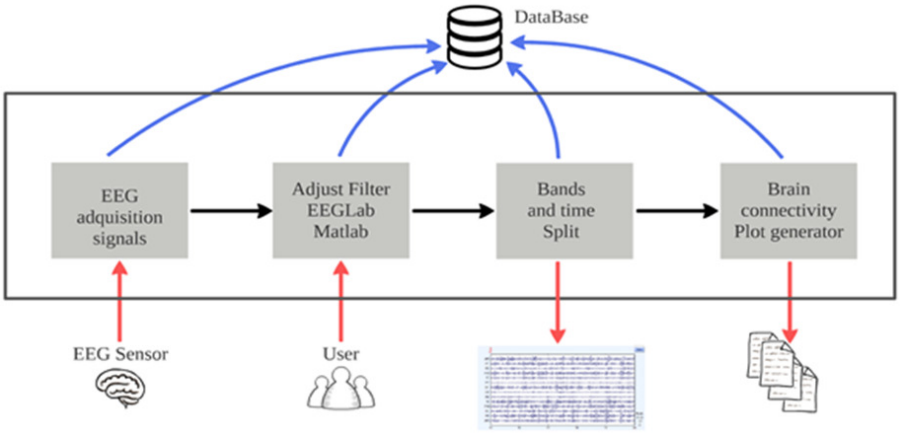


Figure 3: System used to obtain neurocognitive patterns and brain dominances

Correlation and Synchrony Analysis. To estimate inter-channel connectivity, Pearson and Spearman correlation coefficients were computed between all possible electrode pairs, as shown in Figure 4. The resulting 14×14 correlation matrices were visualized as cross-correlation maps, representing functional synchrony between cortical regions. A decision-making tool was implemented in MATLAB® using Rapid Application Development (RAD) methodology. This software automatically computed, filtered, and visualized correlation "dominances"—clusters of significant inter-channel correlations—enabling classification of participants by connectivity profiles and comparison with the cognitive style models of Kolb, VAK, and Herrmann [14],[20],[21].

Hilbert–Huang Transform (HHT) Analysis. Nonlinear and non-stationary properties of EEG signals were explored using the Hilbert–Huang Transform (HHT). Each EEG signal was decomposed into Intrinsic Mode Functions (IMFs) through Empirical Mode Decomposition (EMD) [6],[12]. The Hilbert Transform was then applied to each IMF to compute instantaneous amplitude and frequency,

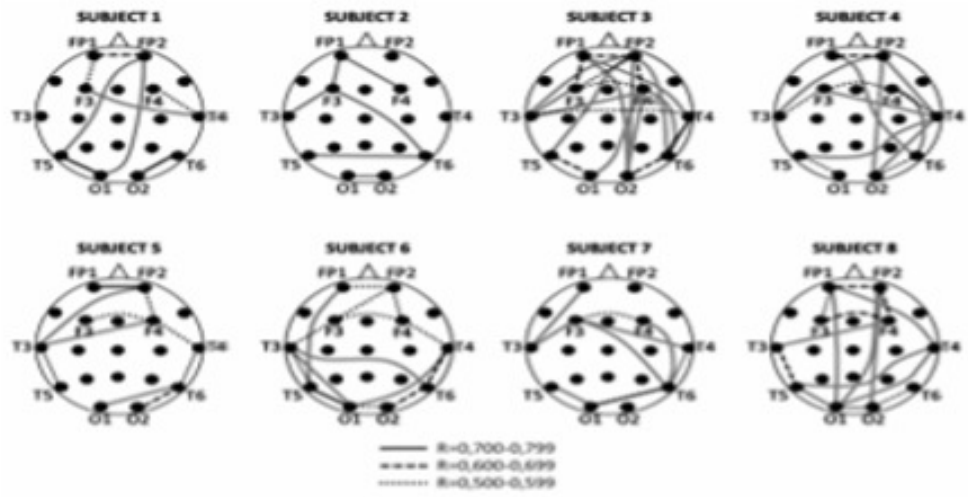


Figure 4: Spearman R cross-correlation map

providing a time–frequency representation of cerebral activity. An example of the decomposition of an EEG signal into IMF components is shown in Figure 5.

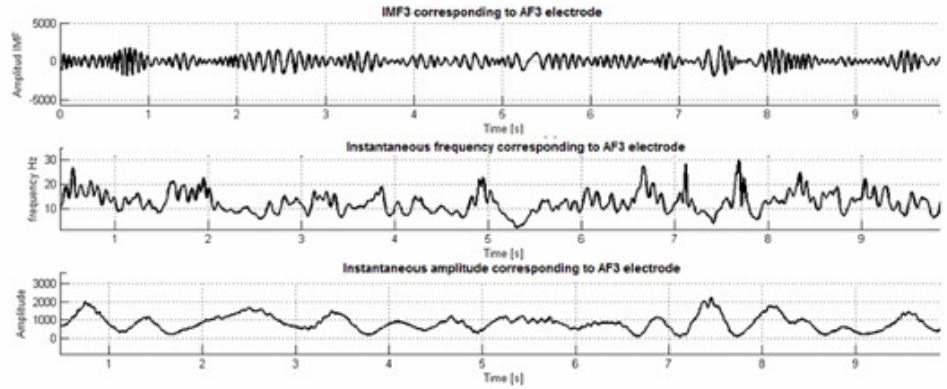


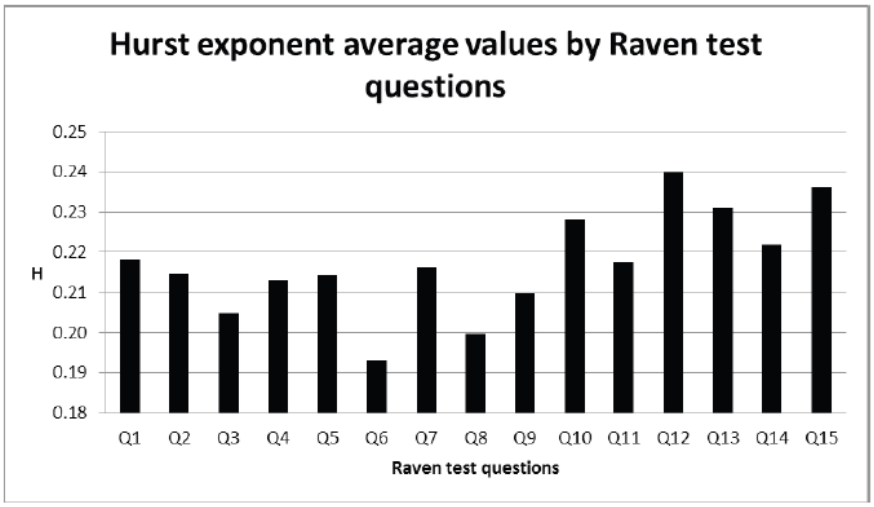
Figure 5: Raw EEG signal decomposition on Channel I (AF3) in IMF3

This method allowed detection of subtle, transient oscillatory components that traditional Fourier-based techniques fail to capture. The analysis provided topographic visualization of energy distributions and frequency shifts across channels, revealing localized activity patterns related to cognitive effort and problem-solving stages.

Hurst Exponent and Order–Chaos Analysis. The Hurst exponent (H) was used as an estimator of long-memory behavior and persistence in EEG signals. Values of $H > 0.5$ indicated self-similar and persistent dynamics, whereas $H < 0.5$ suggested anti-persistence or more irregular dynamics (Figure 6). The exponent was calculated at multiple time scales (1, 3, 12, 30, and 120 s) to assess the temporal variability of brain order and chaos. Inter-hemispheric asymmetries were analyzed across frontal, temporal, and occipital regions, allowing the evaluation of functional dominance during mental projection and resting-state conditions. Synchrony between order–chaos fluctuations was quantified using linear correlations between H time series across electrode pairs [10],[11],[21].

As a result, handling EEG information can be understood as a process that starts from the original recordings and successively generates new data sets, each of which depends on both the data used and the procedures, methods, and parameters applied. This process presents an information management problem that goes beyond just storing the original signals. It is necessary to preserve not only the data but also the context that allows determining how the different results were generated and how they relate to each other.

Multidimensional Data Management (OLAP Model). Online Analytical Processing (OLAP)-



Hurst Exponent variation according execution of Raven test. Questions 1 to 15 (Q1 to q15). Whole brain data.

Figure 6: Hurst Exponent in Raven Test

based model was designed to integrate EEG signals, derived metrics (correlation, HHT, Hurst), and contextual metadata (participant profiles, task conditions, preprocessing methods) [16],[17].

The model was implemented as a Relational OLAP (R-OLAP) structure with a specialized multidimensional query language (MDQL). A middleware layer translated user queries into SQL operations, providing multidimensional operators such as *slice*, *dice*, *roll-up*, and *drill-down* directly over the EEG data cube. This structure ensured semantic traceability between raw EEG data, analytical outputs, and derived visualizations, supporting longitudinal comparison across experiments and participants. The database design followed neuroinformatic standards to preserve data integrity, confidentiality, and interoperability.

Data and metadata organization can be carried out through conceptual models based on the entity-relationship (E/R) model and relational structures. These data organizations provide a suitable representation of the experimental process and allow preserving both the original records and the information needed to describe the conditions in which they were acquired. Their main focus, however, has traditionally been on the acquisition, documentation, and preservation of RAW data.

Original EEG data storage-oriented models generally consider five main components: the characterization of the experiment; the properties of the experimental subject; the features of the instrument used; and the processing and analytical method applied to the data. This organization provides an adequate foundation for documenting the studies carried out and allows for reconstructing the conditions under which the records were obtained.

Data-oriented E/R models for RAW data have several strengths. First, they offer high traceability of the acquisition process, making it possible to reconstruct the capture sequence and the experimental elements involved. They also provide relatively stable structures consistent with common practices in scientific database design. Another positive feature is their extensibility, as they allow for the incorporation of new descriptive entities when necessary. Likewise, the conceptual separation between subject, experiment, instrument, and signal provides semantic clarity for representing the experimental context.

These features make existing models suitable for answering questions related to the origin and acquisition conditions of the data. However, when the focus shifts from preserving the original signals to managing and analyzing the results derived from them, new needs start to emerge that these structures can't always adequately represent.

2.3.3 Limitations for the representation

The evolution of EEG analysis methods has greatly increased the amount and variety of information generated. Nowadays, advanced spectral techniques, nonlinear methods, adaptive decompositions, functional connectivity analysis, and various temporal and spatial analysis procedures can be used. In this context, models mainly designed to document acquisition have several limitations.

A first difficulty relates to the representation of processing and analysis methods. In many cases, existing models do not allow for a systematic description of the method applied at each stage, its parameters, versions, input data, and generated results. The absence of this information makes it difficult to reconstruct the analytical process and limits reproducibility and methodological auditing. A second difficulty relates to the heterogeneity of the results. EEG processing can produce dynamic correlation matrices, spectra, indices calculated for different scales, complexity time series, adaptive decompositions, topographic representations, and numerous other structures. These results can have different dimensionality and varying levels of temporal or spatial granularity.

There is also a limitation when it comes to performing aggregation operations and analytical navigation. Studying the results might require summarizing information by brain regions, comparing groups of subjects, aggregating time intervals, or moving between different levels of a certain hierarchy. Models mainly designed for transactional operations aren't necessarily built to efficiently handle this type of query. Another challenge is scalability when dealing with dense results. Certain methods produce large volumes of information, like correlations calculated over multiple time windows or dynamic connectivity calculated between pairs of channels. Representing this with conventional relational structures can introduce redundancy, increase index complexity, and affect response times.

Finally, there is a problem with semantic integration. Each analytical method can introduce its own concepts, scales, and levels of detail. To integrate their results, it is necessary to relate concepts such as experiment, subject, channel, brain region, time interval, scale, method used, and type of result obtained. These limitations indicate that managing the derived results requires a different perspective than the one used solely to document signal acquisition.

3 Multidimensional nature of EEG information

The variety of information involved in an EEG experiment allows you to identify multiple perspectives from which both the original signals and the results obtained later can be studied. Some of these perspectives correspond to physically identifiable elements, like the instruments used or the participating subjects. Others are related to less tangible concepts, like time, the activity performed by the subjects, or the methods used to process and analyze the recordings. Each of these perspectives provides a different context for interpreting the information and can be considered a dimension of analysis.

3.1 Dimensions associated with the experimental and analytical process

Among the main dimensions shown in Figure 7, the following can be identified:

- **Activity.** This represents the activity or experimental condition in which the subject is engaged during EEG acquisition. It allows the observed signal characteristics and derived results to be linked to the tasks or conditions defined in the experimental protocol.
- **Instrument.** This represents the equipment used during acquisition. It includes the technical features necessary to understand the instrumental conditions under which the signals were obtained.
- **Subject.** This describes the characteristics of the experiment participants. Depending on the study's objectives, these characteristics may include physical, physiological, psychological, emotional, or other relevant background variables.

- **Time.** It represents the temporal information needed to locate both the records and the results obtained from them. This dimension can involve different levels of granularity depending on the analytical procedure used.
- **Method.** It represents the technique, procedure, or method used to process or analyze the EEG records. Considering the diversity of available techniques, it is necessary to link each result to the method by which it was generated and to the relevant characteristics of its application.

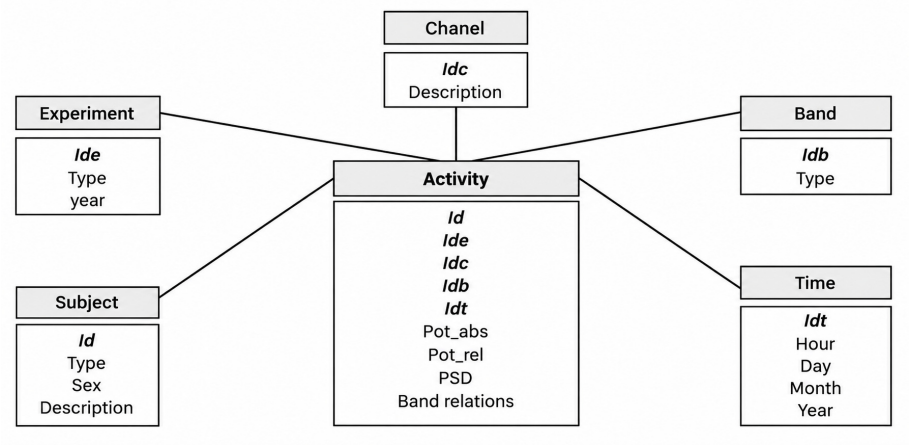


Figure 7: Dimensions Subject, Experiment, Channel, Band and Time

3.2 Dimensions, granularity, and context of the results

The multidimensional nature becomes more evident when considering the variety of results produced by analytical methods.

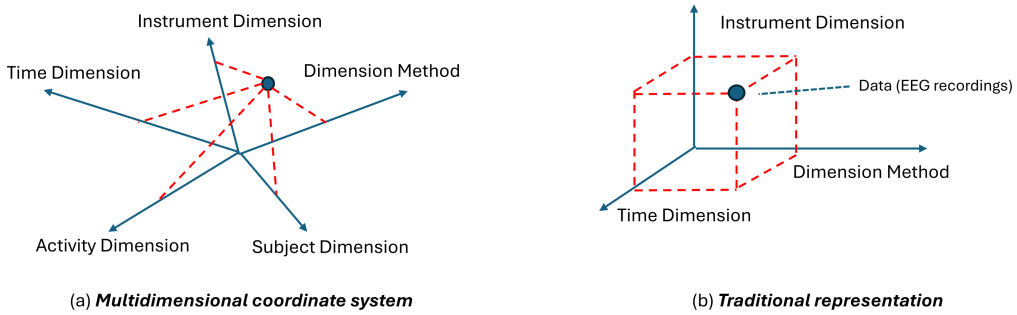
A result can correspond to an entire subject, a specific channel, or a relationship between two channels. It can represent a complete experiment, an experimental stage, a time interval, or a time window. Similarly, it can be associated with a specific spectral band, analysis scale, brain region, or experimental condition. Therefore, the dimensions don't just serve as descriptive mechanisms. They also set different levels from which information can be organized, accessed, and compared. This allows interpreting the results through a conceptual space defined by different coordinates. For example, an analysis can simultaneously consider a specific method, a set of subjects, a type of instrument, and a time sequence. Another analysis might focus on the relationship between certain brain regions under different experimental conditions. The possible combinations depend on the phenomenon being studied and the method applied. In this sense, the information generated has an inherently multidimensional structure.

Multidimensional analytical structures, including those traditionally used in OLAP environments, provide useful concepts for representing this type of organization through dimensions, hierarchies, levels of granularity, measures, and aggregation mechanisms. However, the nature of the EEG problem also introduces requirements related to the heterogeneity of the results and the need to preserve the relationship between the different stages of the analytical process. For this reason, multidimensionality should be understood as a fundamental property of information that needs to be considered in management contexts, and not just as a technological choice for its implementation.

3.3 Need for an integrated context for outcomes derived from EEG

Previous background shows an important difference between two related but conceptually distinct problems. The first one is about managing the original EEG recording and its experimental context. For this purpose, models focused on RAW signals provide suitable structures for representing subjects, experiments, instruments, and acquired data. The second one is about managing the information generated through processing and analyzing these signals. This problem includes new elements: different

methods, multiple processing stages, parameters, intermediate and final results, different spatial and temporal granularities, large volumes of data, and dependency relationships between the results obtained.



(a) ***Multidimensional coordinate system*** (b) ***Traditional representation***

Figure 8: Multidimensional representation of EEG data

The use of separate files, directory structures, or tables specifically designed for each procedure allows for meeting particular needs, but makes it difficult to have an integrated view of the analytical process. This can limit systematic comparison between subjects or experiments, the relationship between the methods used and their results, and the reconstruction of the different stages through which certain information was obtained. As a result, a context is required that can integrate experimental information with the processes applied to the signals and with the results produced by these processes.

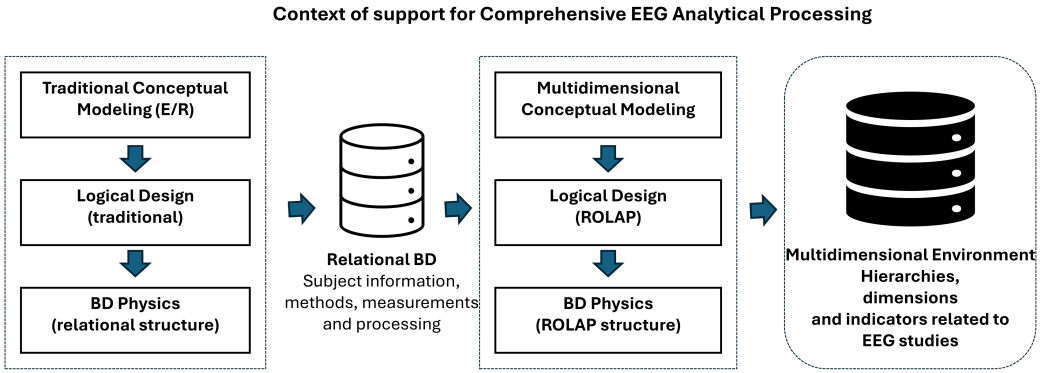


Figure 9: Analytical Framework for Integral EEG Processing

Such a context should at least allow for:

- representing the experimental conditions associated with the source data;
- identifying the data used as input in each process;
- representing the applied methods and their main characteristics;
- keeping the parameters necessary to interpret the execution of the methods;
- managing heterogeneous results with different levels of granularity;
- representing temporal and spatial dimensions and hierarchies;
- establishing relationships between subjects, experiments, methods, and results;
- facilitating search, comparison, aggregation, and navigation operations on the results;
- maintaining traceability between the original data, processing stages, and the results obtained.

In this way, the problem stops being exclusively about storing EEG records and starts considering the cycle of generating and using the derived information. Defining a context that allows storing, organizing, processing, and managing these results is, therefore, a necessary step to move from systems mainly focused on preserving signals to environments capable of systematically supporting EEG information analysis. Based on these requirements, the next section presents a context designed to organize the results generated through different processing and analysis procedures, preserving both their multidimensional characteristics and the relationships that allow reconstructing their origin and use.

4 Integrated architecture for EEG information and knowledge management

The growing diversity of methods used for processing and analyzing electroencephalographic signals generates sets of information that show significant differences in structure, dimensionality, granularity, and meaning. Along with the signals originally recorded during experiments, the analytical processes produce scalar indicators, transformed time series, connectivity matrices, spectral representations, and other derived structures. Integrated management of these elements therefore requires considering not only the original recordings but also the procedures through which they were processed, the experimental context in which they were obtained, and the results generated during the successive stages of analysis. To address this need, a proposed integrated architecture is made up of three main components: a Primary EEG Data Repository, a Semantic Add-on, and a Multidimensional Environment. On top of these components, there is a cross-cutting Knowledge Management layer supported by Artificial Intelligence, aimed at jointly leveraging the primary information, the semantic context, and the accumulated analytical results.

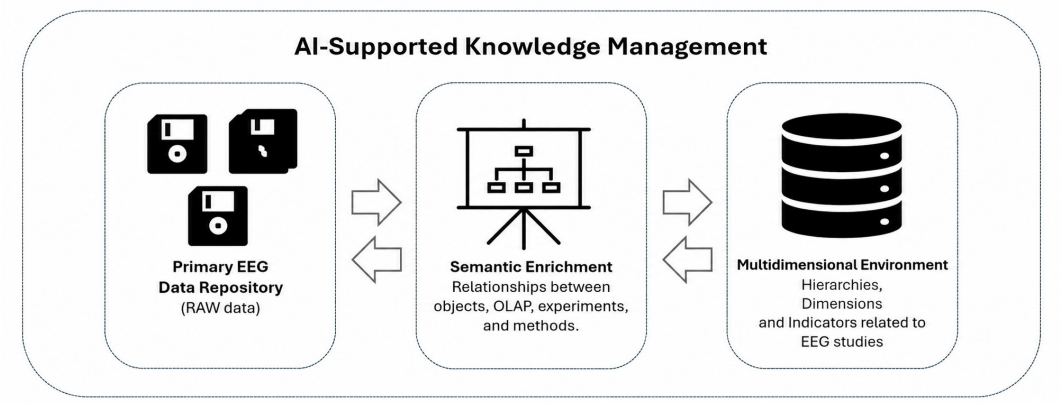


Figure 10: General context of the AI-supported knowledge management platform

The separation between the three components is due to the different nature of the information being managed. The primary data repository stores the original experimental evidence; the multidimensional environment organizes the products derived from processing to make their analytical use easier; and the semantic complement establishes the relationships needed to reconstruct the context and origin of those results. The artificial intelligence layer uses this integrated structure to identify patterns among results from different experiments and support the gradual accumulation of knowledge. Figure 10 shows the overall structure of the proposal. The relationships between its components represent logical integration and interoperability and don't necessarily imply physical transfer or duplication of data. This distinction allows each component to use storage structures suited to the type of information it manages, while at the same time maintaining an integrated view of the experimental and analytical process.

In general terms, the process can be represented through the following conceptual sequence:

$$\textit{Primary data} \rightarrow \textit{Experiment} \rightarrow \textit{Method} \rightarrow \textit{Result} \rightarrow \textit{Multidimensional Object}$$

This sequence constitutes the analytical lineage of the result. From an indicator stored in the multidimensional environment, it should be possible to identify the method by which it was generated, the experiment it comes from, and the primary records that made up its source. The relationship can also be traced in the opposite direction. From a given set of original EEG data, it is possible to identify the methods that have been applied, the products obtained, and the analytical structures in which they are available. The semantic complement thus takes on an integrating role within the architecture. Its representation allows linking objects that correspond to different storage paradigms without requiring that all of them be incorporated into a single data structure. This separation also favors the extensibility of the model. New types of methods, processes, or relationships can be incorporated into the semantic complement without necessarily having to modify the physical organization of primary data or multidimensional objects.

4.1 Primary EEG data repository

The Primary EEG Data Repository is the starting point for the information managed by the architecture. Its role is to preserve the signals obtained directly during experiments, along with the elements needed for their identification and later retrieval. This separation is important because primary data have different characteristics from derived results. While the former are high-resolution records obtained directly with acquisition instruments, the latter are transformations or indicators produced through specific analytical procedures.

The architecture considers that RAW data should be preserved regardless of the results generated later. This way, the same set of original data can be subjected to different processing methods as available techniques evolve or new research questions arise. This feature prevents analytical results from becoming irreversible representations of the experiments. When new patterns are identified or new methods are introduced, the researcher can go back to the primary data and generate new derived products without altering the original information. The repository thus acts as a persistent source of experimental evidence and provides the necessary basis for maintaining the reproducibility of different analysis processes.

4.2 Semantic complement

Primary data and analytical results have limited value if they are detached from the context in which they were obtained. To address this problem, a component called Semantic Complement is incorporated, responsible for representing the relationships between the main elements involved in the experimental and analytical cycle. This component allows linking projects with experiments and their context, connecting experiments with the primary data generated during their execution, representing the methods applied during processing, and establishing relationships with the results produced and the multidimensional objects in which they are stored.

Its role is not to duplicate the measurements contained in the multidimensional environment or the signals stored in the primary repository. Its goal is to maintain the meaning, context, and origin of these elements.

4.3 Multidimensional Environment

The Multidimensional Environment is focused on storing and making use of results obtained through the processing and analysis of EEG signals. Its organizational principle is different from that used for primary recordings. Multidimensional objects are not structured based on a particular experiment, but mainly according to the nature of the produced results, their dimensions, and their granularity. Consequently, different experiments can contribute information to the same type of multidimensional object when they generate analytically compatible results. For example, connectivity results obtained in different experiments could be integrated into a common multidimensional structure when they share dimensions such as subject, experimental condition, time, channel, or brain region and use equivalent measures.

Similarly, specific structures could be defined for spectral results, causality measures, complexity indices, or other families of results. This organization allows for a conceptual distinction between

the experiment and the analytical object. The experiment provides the context in which certain data were obtained, while the multidimensional object organizes comparable results regardless of their experimental origin. This feature supports cross-sectional analysis. Results produced by independent projects can be compared when they have compatible analytical meanings, expanding the possibilities for reusing accumulated information and expert learning.

Architecture does not assume the existence of a fixed number of cubes. New types of results may require new multidimensional objects, while results compatible with existing structures can be incorporated into them. This creates an extensible environment whose evolution mainly depends on the types of analytical information that need to be represented.

The multidimensional model also provides suitable mechanisms for working with dimensions, hierarchies, and different levels of aggregation. These features make it possible to analyze results from different perspectives, for example, moving from channels to brain regions, from individual subjects to groups, or from short time windows to longer periods.

Multidimensionality had previously been identified as an intrinsic feature of EEG-derived data, whose temporal, spatial, experimental, and methodological perspectives are not adequately represented through flat structures. In the proposed architecture, this feature is used as a basis to commonly organize results from different experiments.

5 Knowledge management supported by artificial intelligence

The three components described allow for maintaining primary data, contextual information, and analytical results within an integrated structure. However, the gradual accumulation of results from multiple experiments opens up an additional possibility: using artificial intelligence techniques to identify patterns that were not necessarily defined during the original study design. The Knowledge Management layer supported by Artificial Intelligence is proposed for this purpose. Its function is not only to automate queries on existing databases. The goal is to use quantitative results and semantic relationships together to discover patterns, contextualize them, and generate knowledge that can later be examined and reused by researchers.

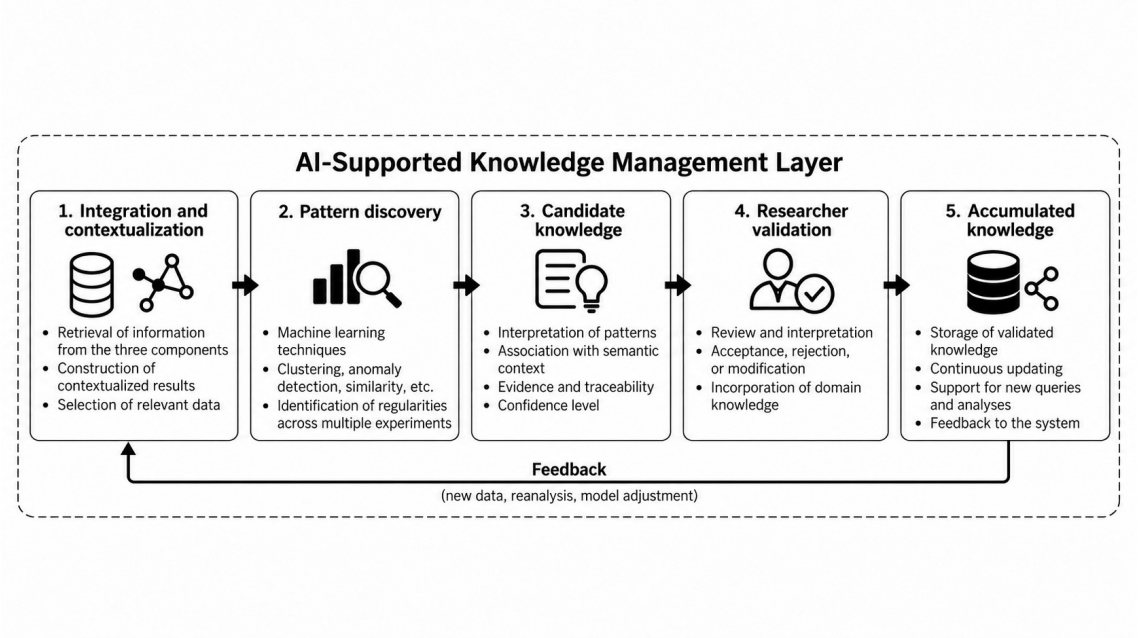


Figure 11: Knowledge management layer supported by AI

The architecture of this layer is deliberately kept independent of specific algorithms. The heterogeneous nature of EEG results makes it unlikely that a single artificial intelligence technique would be suitable for all multidimensional objects. For this reason, a functional structure is proposed in which different learning methods can be incorporated depending on the characteristics of the prob-

lem being analyzed. Its operation can be summarized in four processes: information integration and contextualization, pattern discovery, candidate knowledge generation, and knowledge validation and accumulation.

5.1 Integration and Contextualization

The first stage aims to build suitable sets of information for discovery processes. The indicators contained in the multidimensional environment provide quantitative information, but interpreting them requires knowing the conditions under which they were obtained. The semantic complement allows them to be connected with experiments, methods, and other relevant elements, and, when necessary, to access the primary EEG data again. Therefore, the unit used by artificial intelligence processes shouldn't just be a standalone measurement, but a contextualized result.

The context can include, depending on the analysis, information related to the experiment, experimental condition, subject or group characteristics, the method used, and other elements represented in the semantic complement.

Access to RAW data can be included when the analytical procedure requires it. However, it's not necessary to reprocess the original signals in every discovery operation. In many cases, the indicators previously stored in the multidimensional environment will provide the necessary quantitative information. This distinction helps limit computational costs and use the primary data when it adds extra information or when new procedures need to be applied.

5.2 Pattern Discovery

Once the sets of contextualized results have been built, machine learning techniques can be applied to identify regularities present in the information. The architecture initially considers methods that do not require having predefined categories. Clustering techniques can be used to recognize results with similar behaviors; anomaly detection methods can identify infrequent configurations; dimensionality reduction procedures can reveal underlying structures; and similarity analyses can help establish relationships between experiments or sets of results. The goal of this stage is not to replace scientific analysis or to automatically attribute physiological meaning to the patterns found. Its purpose is to reduce the search space and highlight potentially relevant regularities within increasing volumes of results.

A central aspect of the proposal is that these patterns can involve results coming from multiple experiments. For example, imagine that different studies have generated connectivity indicators stored in the same multidimensional object. A clustering procedure could identify sets of results with similar behavior without initially considering which project they come from. Using the semantic component, it could later be determined whether these results share certain experimental conditions, processing methods, or other characteristics. The combination of quantitative regularity and semantic context thus provides a basis for moving from identifying statistical structures to interpreting them (candidate knowledge).

5.3 Generation of candidate knowledge

The patterns detected using artificial intelligence techniques should not be directly interpreted as validated scientific knowledge. The system suggests distinguishing between detected patterns, candidate knowledge, and validated knowledge.

$$\textit{Pattern detected} \rightarrow \textit{Knowledge candidate} \rightarrow \textit{Knowledge validated}$$

A detected pattern corresponds to a regularity identified computationally. This pattern becomes candidate knowledge when it can be related to a context and a recognizable set of evidence. For example, the recurring existence of certain connectivity relationships could become candidate knowledge when it is determined that the behavior appears in different experiments associated with comparable conditions. Candidate knowledge should retain enough references to reconstruct the evidence that

allowed it to be generated. This includes the analytical results involved, the corresponding experiments, the methods applied, and the artificial intelligence procedure through which it was identified. This mechanism maintains an explicit separation between algorithmic discovery and scientific interpretation and prevents an automatically produced association from being incorporated as knowledge without additional evaluation.

5.4 Validation and accumulation of knowledge

Validation by experts is a necessary stage of the process. Algorithms can identify mathematical or statistical structures, but their scientific relevance must be evaluated considering domain knowledge, the characteristics of the experiments, and the strength of the available evidence. Researchers can accept, reject, or reformulate the candidate knowledge and eventually include observations that define the conditions under which it is applicable. Knowledge that passes this process can be added to a repository meant to progressively preserve the patterns considered relevant. However, the stored knowledge should not be seen as definitive. New experiments can bring additional evidence capable of confirming, modifying, or weakening previously identified patterns. The process therefore has a second direction:

$$\textit{Existing knowledge} + \textit{New results} \rightarrow \textit{Reevaluation} \rightarrow \textit{Updated knowledge}$$

This feature turns the knowledge repository into an evolving structure that grows along with the available experimental information.

5.5 Functional architecture of the artificial intelligence layer

Functional architecture distinguishes two complementary cycles. The first corresponds to the discovery cycle, through which accumulated information is analyzed in order to identify new patterns. This cycle can run when new results are added, when the amount of available information significantly increases, or when researchers need to explore a specific set of indicators. The second corresponds to the knowledge reuse cycle. Previously validated patterns can be used as background to interpret new results, select relevant information sets, or raise new analysis questions. In this way, artificial intelligence doesn't just act on the data available at a given moment, but can progressively benefit from the experience accumulated by the system and the validation carried out by researchers. Figure 11 summarizes the proposed operation for the artificial intelligence layer.

The future use of supervised techniques is a possible evolution of this process. Initially, when there is little labeled information, unsupervised techniques may predominate. As specialists evaluate patterns and candidate knowledge, the decisions made can become new sources of information for supervised learning procedures. The architecture thus supports different levels of maturity without relying on any particular technology.

5.6 Interaction with Language Models

Large-scale language models can complement the architecture by facilitating interaction between researchers and the accumulated information. However, their role should be distinguished from the mechanisms responsible for discovering quantitative patterns.

In the proposal, a language model can act as an interface to formulate queries, locate relevant information, and present integrated explanations. For example, a researcher could ask what connectivity behaviors have been observed repeatedly under certain experimental conditions. The answer would first require identifying the relevant experiments through the semantic complement, retrieving the corresponding results from the multidimensional environment, and determining whether there are any patterns or previously validated knowledge related to the query.

The language model could later organize this information and present an understandable explanation, keeping the reference to the evidence used. This setup is compatible with retrieval-augmented generation (RAG) mechanisms, where the model receives information retrieved directly from the architecture's components as context before producing a response. Therefore, the language model wouldn't

be the primary source of scientific knowledge, but rather a tool for interacting with, integrating, and explaining information managed by the system itself. This distinction is important to maintain traceability and limit the generation of statements that can't be linked to available experimental evidence.

5.7 Traceability of Generated Knowledge

The integration of the three components makes it possible to extend the traditional concept of traceability from experimental data to the knowledge generated. Each candidate or validated piece of knowledge should be able to be linked to the patterns that originated it and, through them, to the analytical results, methods, and corresponding experiments. Conceptually:

$$Knowledge \rightarrow Pattern \rightarrow Results \rightarrow Methods \rightarrow Experiments \rightarrow Primary\ Data$$

This chaining provides a relevant feature for systems focused on scientific research: any knowledge generated through artificial intelligence procedures can be linked back to the evidence that made it possible. Traceability also supports reproducibility. In addition to identifying the data used, there should be enough information preserved about the analytical procedures and AI methods applied during the discovery. In this way, the incorporation of AI is not a black box layered over the data, but rather a set of processes integrated within the same framework of provenance and information management.

6 Conclusions

EEG analysis generates results with different temporal, spatial, and methodological granularities. Managing these results requires more than preserving the original recordings: their meaning also depends on the experimental conditions, processing decisions, and methods through which they were obtained. This paper proposes an integrated architecture that connects a repository of primary EEG data, a semantic component for experimental and analytical provenance, and a multidimensional environment for organizing derived results.

The central contribution is to organize analytically compatible results across experiments while preserving their links to the source recordings and the procedures that produced them. This approach provides a basis for comparing indicators and exploring them at different levels of detail without losing the context needed to interpret them. The proposed AI-supported layer extends the architecture toward the discovery of patterns and the generation of candidate knowledge, subject to examination and validation by researchers.

The conceptual proposal of frame and multidimensional architecture supported by AI Knowledge management opens up an important design space. Future work should implement it with data from multiple EEG studies and evaluate the compatibility of their results, the usefulness of multidimensional queries, the reconstruction of analytical provenance, and the quality of the knowledge candidates generated with AI support.

Acknowledgements

The authors would like to acknowledge the support of the Facultad de Ingeniería and the Vicerrectoría de Investigación y Doctorados at Universidad San Sebastián, Chile, as well as the Departamento de Matemática y Ciencia de la Computación of the Facultad de Ciencia at Universidad de Santiago de Chile, for the institutional support provided during this research project.

Funding

This work is funded as part of the authors' academic activities.

Author contributions

The authors contributed equally to this work.

Conflict of interest

The authors declare no conflict of interest.

References

[1] Abdul Rashid, N., Taib, M.N, Lias, S., Sulaiman, N., Murat Z. H., Abdul Kadir, R.S., Learning Style Classification related to IQ and Stress based on EEG., *Social and Behavioral Sciences*, 29 (2011), p. 1061-1070.

[2] Amunts, K.; Ebell, C.; Muller, J.; Telefont, M.; Knoll, A.; Lippert, T. (2016). The Human Brain Project: Creating a European Research Infrastructure to Decode the Human Brain, *Neuron*, 92(3), 574–581, 2016.

[3] Badcock, N.A.; Mousikou, P.; Mahajan, Y.; de Lissa, P.; Thie, J.; McArthur, G. (2013). Validation of the Emotiv EPOC® EEG gaming system for measuring research quality auditory ERPs, *PeerJ*, 1, e38, 2013.

[4] Buzsáki, G. (2006). *Rhythms of the Brain*, Oxford University Press, 2006.

[5] Córdova, F.; Díaz, H.; Cifuentes, F.; Cañete, L.; Palominos, F. (2015). Identifying Problem Solving Strategies for Learning Styles in Engineering Students subjected to Intelligence Test and EEG Monitoring, *Procedia Computer Science*, 55, 18–27, 2015.

[6] Córdova, F.M.; Atero, R.; Cifuentes, F. (2017). Brain Topography Method based on Hilbert-Huang Transform, *Procedia Computer Science*, 122, 873–880, 2017.

[7] Córdova, F.; Troncoso, N.; Díaz, H.; Palominos, F.; Cañete, L. (2016). Design of a decision-making tool to identify cross-correlation brain dominances, *Procedia Computer Science*, 91, 65–74, 2016.

[8] Delorme, A.; Makeig, S. (2004). EEGLAB: an open-source toolbox for analysis of single-trial EEG dynamics, *Journal of Neuroscience Methods*, 134(1), 9–21, 2004.

[9] Díaz, H.; Maureira, F.; Córdova, F. (2017). Temporal scaling and inter-individual hemispheric asymmetry of chaos estimation from EEG time series, *Procedia Computer Science*, 122, 339–345, 2017.

[10] Diaz, H.; Córdova, F.; Cañete, L.; Palominos, F.; Cifuentes, F.; Sánchez, C.; Herrera, M. (2015). Order and chaos in the brain: fractal time series analysis of the EEG activity during a cognitive problem solving task, *Procedia Computer Science*, 55, 1410–1419, 2015.

[11] Díaz, H.; Maureira, F.; Flores, G.; Fuente, I.; García, F.; Maerten, P.; Muñoz, S.; Parra, F.; Córdova, F. (2018). Moving correlations and chaos in the brain during closed eyes basal conditions, *Procedia Computer Science*, 139, 473–480, 2018.

[12] Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.C.; Tung, C.C.; Liu, H.H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proceedings of the Royal Society A*, 454, 903–995, 1998.

[13] Khashman, A. (2010). Modeling Cognitive and emotional processes: A Novel Neural Network Architecture, *Neural Networks*, 23(10), 1155–1163, 2010.

[14] Kolb, D. (1984). *Experiential Learning: Experience as the Source of Learning and Development*, Prentice Hall, 1984.

[15] Kostas, D., Aroca-Ouellette, S., and Rudzicz, F. (2021). BENDR: Using Transformers and a Contrastive Self-Supervised Learning Task to Learn From Massive Amounts of EEG Data. *Frontiers in Human Neuroscience*, 15, 653659. https://doi.org/10.3389/fnhum.2021.653659

[16] Palominos, F.; Díaz, H.; Córdova, F.; Cañete, L.; Durán, C. (2017). A Solution for Problems in the Organization, Storage and Processing of Large Data Banks of Physiological Variables, *International Journal of Computers Communications & Control*, 12(2), 276–290, 2017.

[17] Palominos, F.; Córdova, F.; Durán, C.; Nuñez, B. (2020). A Simpler and Semantic Multidimensional Database Query Language to Facilitate Access to Information in Decision-making, *International Journal of Computers, Communications & Control*, 15(4), Article 3900, 2020.

[18] Pfurtscheller, G.; Lopes da Silva, F.H. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles, *Clinical Neurophysiology*, 110(11), 1842–1857, 1999.

[19] Popivanov, D.; Mineva, A.; Krekule, I.; Stomonyakov, V. (2006). EEG patterns in cognitive tasks: persistence and scaling properties, *Physiological Measurement*, 27(6), 571–587, 2006.

[20] Rashid, N.A.; Taib, M.N.; Lias, S.; Sulaiman, N. (2010). Classification of learning style based on Kolb's Learning Style Inventory and EEG using cluster analysis approach, *Engineering Education (ICEED), 2nd International Congress on*, 64–68, 2010.

[21] Raven, J. (2000). The Raven's progressive matrices: change and stability over culture and time, *Cognitive Psychology*, 41, 1–48, 2000.

[22] Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. H., and Faubert, J. (2019). Deep learning-based electroencephalography analysis: a systematic review. *Journal of Neural Engineering*, 16(5), 051001. https://doi.org/10.1088/1741-2552/ab260c

[23] Smith, S.J.M. (2005). EEG in the diagnosis, classification, and management of patients with epilepsy, *Journal of Neurology, Neurosurgery, and Psychiatry*, 76(Suppl 2), ii2–ii7, 2005.

[24] Song, T., Zheng, W., Song, P., and Cui, Z. (2020). EEG emotion recognition using dynamical graph convolutional neural networks. *IEEE Transactions on Affective Computing*, 11(3), 532–541. https://doi.org/10.1109/TAFFC.2018.2817622

[25] Theiler, J.; Eubank, S.; Longtin, A.; Galdrikian, B.; Farmer, J.D. (1992). Testing for nonlinearity in time series: The method of surrogate data, *Physica D: Nonlinear Phenomena*, 58(1–4), 77–94, 1992.

[26] Tononi, G.; Edelman, G.M. (1998). Consciousness and complexity, *Science*, 282(5395), 1846–1851, 1998.

[27] Zhou, C.; Zemanová, L.; Zamora, G.; Hilgetag, C.C.; Kurths, J. (2006). Hierarchical organization unveiled by functional connectivity in complex brain networks, *Physical Review Letters*, 97(23), 238103, 2006.



Copyright ©2026 by the authors. Licensee Agora University, Oradea, Romania.
This is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International License.
Journal's webpage: http://univagora.ro/jour/index.php/ijccc/



This journal is a member of, and subscribes to the principles of,
the Committee on Publication Ethics (COPE).
https://publicationethics.org/members/international-journal-computers-communications-and-control

Cite this paper as:
Córdova, F.; Palominos, F. (2026). Contextualizing EEG-Derived Results: A Semantic and Multidimensional Architecture for Cross-Study Analysis, *International Journal of Computers Communications & Control*, 21(6), 7525, 2026.
https://doi.org/10.15837/ijccc.2026.6.7525