The evolution of a standardized database via the availability of tools

As I understood things, the database that was originally conceptualized was to be one where the information was origin independent. The idea being that subsets of the database could be directly compared and contrasted even though the data within these subsets may have been collected at different labs. This implies a reasonable amount of standardization.

The current state of research, at least in the area of human brain electrophysiology, is relatively anarchistic in that other than a loose adherence to the 10-20 placement system, each lab essentially goes its own way. Because of this, much of the variability in the data obtained for different studies and at different labs may be due to factors unrelated to a variable that one might want to consider when comparing subsets within a database. This is why, although useful as a reference source, an archival database of existing data would not be adequate. However, if we consider an origin independent database as a future goal we can begin to plan on ways to consolidate methodologies and experimental parameters in order to create sufficient standardization to make such a database scientifically feasible.

Many sources of variability are those that are not intrinsic to the data itself and therefore would not pose a problem to the database as long as the raw data and sufficient description of this data (or metadata) was available. For example, the database might store the results of a standard set of analyses applied to the data of each study that contributed to the database, however, any researcher would be free to analyze the raw data with whatever analyses they saw fit. The development of a standard set of analyses would be facilitated by providing the software and instruction in the use of this software to those who provide data to the database. Often types and/or characteristics of analyses done depend on the analyses software package that is available to the researcher. If a standard analyses package is made readily available as part of database participation, researchers will only go through the additional effort to do different analyses if they deem there to be sufficient scientific rational.

If a researcher develops software for a new analysis, that researcher can provide a website that links to the database that provides the analysis software and instructions on its use. If that analysis provides an improvement to the standard analyses researchers will seek it out. If a number of researchers begin to use this analysis it can become incorporated into the standard package and existing data in the database can be reanalyzed with this method. If, on the other hand, this alternative analysis does not provide any improvement to the existing package researchers will not be motivated to use it. In this way the evolution of the database can be driven by what are determined to be scientifically necessary analyses while eliminating variability in analyses characteristics due to competing variants of analysis packages. The only critical aspect of this is whether or not new analyses will require data in addition to what was initially required.

Similarly, there are differences in the treatment of data that can be removed by mathematical transformations. As a simple example, the polarity of EEG data in the database could be arbitrarily decided. Users could determine their visualization preferences and the data could be converted online to reflect these preferences. Likewise, data could be mathematically re-referenced given that the data from the necessary electrodes for that re-referencing was initially collected.
This leads us to the true challenge of any proposed database; deciding on a way to standardize those aspects the data that are intrinsic. For example, once EEG data is collected we can not change the location of electrodes, we can not change the stimuli and the rate they were presented, we can not add a response requirement or increase the sampling rate etc.

Although many of these parameters can achieve a good amount of standardization within a laboratory or within a collaborative project, most studies usually differ on a variety of parameters. Unfortunately, differences in these parameters can act as confounding variables when making comparisons across different studies (i.e. subsets of a database). In order for a database to work it is these parameters that must be standardized.

In terms of human nature any attempt to decree a set of standards at the onset would be met with resistance. Most researchers would view imposed standards as restrictions to their intellectual freedom. However, as with analyses, data collection standards could evolve via the availability of experimental protocols and data collection software.

In most experiments, the selection of values for parameters extraneous to the hypothesis are driven by the characteristics of the equipment, software or stimuli and their availability. For example, electrode placement is determined by whether one has a geodesic net or an electrode cap, filter and sampling rate settings are limited to the values allowed by the software, stimuli are obtained from other researchers or are developed within the limitations of the tools available to the researcher etc. In my protocol code stimulus timing is set to multiples of 40ms in order to best match a 75 Hz monitor refresh rate. The fact that a stimulus might be 640 rather than 625 or 650ms is determined solely by the characteristics of the equipment. Similarly, a researcher sampling at 250Hz may be limited to an epoch length of 1024ms because the software can only manage trials of 256 sample points. The creation of an origin independent database could be greatly facilitated if the number of these "necessity dictated" differences could be eliminated.

Quite often in research the experimental protocol does not involve a radically new paradigm but is instead a variation on some other experiment. A researcher will change the parameters that have bearing on an experimental hypothesis while attempting to replicate the remaining parameters in order to maximize the comparability between study results. However, since each researcher is limited by the type of data collection / experiment control / stimulus generation resources on hand, this replication is inexact, thereby decreasing the degree to which two studies actually can be considered comparable. In fact, it could be argued that the reason why seemingly equivalent changes in experimental variables can often produce such different results from different labs is a direct consequence of these differences in methodological parameters.

One simple but powerful way these differences can be minimized is through the availability of resources. Instead of just a general description of how an experiment was done, each study in the database should include the complete set of software tools and stimuli that were used to generate the data from that study. For example, a researcher may develop a protocol to compare ERPs in low versus high vigilance conditions in typical adults. If I'm interested in studying adults with attention deficit in low versus high vigilance conditions I might want to use the same protocol. Usually what would happen is that I would try and develop a similar protocol given my resources. And maybe a third researcher interested in children with AD might develop yet another version of the initial protocol with yet another set of resources.

What we would like to do is use the findings from all 3 studies to gain insight into how vigilance task performance is affected by age and attention deficit, however, the dissimilarities in methodology can potentially confound any comparison. If, for example, the relationship between typical and AD is different in the adult and child studies there is no way we can determine if this
difference is primarily due to age, one or more methodological parameters or their interaction, or
the interaction of AD and age and/or one or more methodological parameters. Usually one of
two things then happens in this situation, the methodological issues are ignored or one or both
labs attempt to address some subset of these issues. For example, I might run another study with
both adults and children using an ISI that is more similar to the ISI used by the 3rd researcher.
That researcher might run another study with children that incorporates conditions that compare
stimulus types. Neither lab might be using an ISI or stimulus type that is exactly the same as the
other and they may both still differ in stimulus duration, analog filter settings etc. Since we still
do not have the same methodology not only do we have the same problems as before, but they
can be compounded if there are effects due to several methodological parameters that interact.
Ultimately, we are left with a few generalizations and many disagreements over specifics such as
exactly which task / cognitive process manipulations effect which components and how (e.g.
peak amplitude or latency)

If, on the other hand, researchers have full access to the protocol and software tools used,
most of these methodological discrepancies can be avoided. While it’s true that the contextual
parameters of an experiment can never be exactly replicated, even within a lab, arguably there is
much greater justification for making comparisons between studies if the experimental
parameters are the same. In other words, to take the vigilance protocol as an example, one would
not expect a rectangular vs. square shaped room or whether the experiment was run by a grad
student vs. post-doc to have much, if any, effect on the results. Conversely, one could easily
hypothesize how the number of trials between rest periods or differences in response requirement
could effect a vigilance task.

One can envision in theory a multidimensional database where each cell contains data
obtained from an experimental condition and where each dimension represents a relevant
parameter. Each adjacent cell would then represent a variation along one dimension. Although it
would not be feasible to develop this database as a uniform matrix, this database (or more
precisely, a set of databases) could evolve as trees of links within the matrix of possible
comparisons. Each experiment would be connected to the database by a link cell condition that
is an exact replication of an existing experimental cell within the database. The remaining
experimental cells of a study would involve steps that each involve a single parameter change. In
this way, every experiment would have a point-of-reference by which it could be compared to
other experiments.

This concept is illustrated below as a 2D graphic. The left-hand panel represents how an
archival database of current research might look if it was overlaid on a matrix of possible
comparisons. The right-hand panel represents how database trees could evolve from unique
paradigms/protocols (designated as black squares) within this matrix. Note that the idea of a
database via availability of experiment tools does not require that every experiment be identical
or be a direct offshoot of an existing paradigm. However, given that a good number of
experiments actually are variants of existing paradigms/protocols, each researcher can maximize
the comparability between experiments if each can start with a protocol identical to a related
experiment and then change a single parameter as a variable of interest rather than having to
recreate with approximations a variety of parameters.
If we can eliminate this parameter variability we can discount most of the potential methodological confounds that might have occurred with across experimental comparisons and essentially have origin independent data. This would allow us to conduct "virtual" experiments with some degree of certainty that our results would be representative of the results obtained in an actual experiment. For example, if the adult typical vs. AD and the child typical vs. AD vigilance studies used the exact same tools for running the experiment and collecting data, the raw data could be combined in order to analyze age x AD interactions. By requiring some overlap between studies we can further assert the validity of virtual comparisons. For example, the child study might include a small N subset of adults as verification that the adult data obtained was comparable to previous adult data.

By continually adding this kind of verification data, different cells in the database will obtain sufficiently large N to begin using the data for normative value comparisons. In addition, large N and adequate metadata will allow post-hoc examination of subgroupings, individual differences and complex interactions that are not possible with the N sizes obtained in single experiments.

Obviously though, in the beginning, not everyone will be able to make use of protocol software even if it is made available because of compatibility problems with existing equipment. However, over time as more protocol tools are made available labs will gravitate to the systems and equipment that are most compatible with the available tools. Eventually there will be a convergence to standardization as various systems are faded out or are restructured to accommodate the emerging standard. One needs to look no further than audio / video formats and computer operating systems to see how simple product availability can evolve into standards. Of course, if standards are derived by pure market forces there is the risk that these standards will reflect the goals of developers rather than the needs and requirements of the users. This can be counteracted by developing collection/presentation/analyses tool software using an open-source code model and general public licensing agreements. In this way, software can be molded to the needs of the community while, at the same time, tracking the developmental history of the software in order to credit and acknowledge the software developers.

As part of the evolution of these standards and subsequent databases, the neuroscience community will need to decide on core requirements and the criteria for division into equivalent groupings. For example, are there a minimum number of electrodes or specific electrode locations that must be included? Are there specific tests that must be used to discriminate clinical populations? The more standards that can be decided a priori the faster the database can evolve and the easier it will be to link the different trees that develop from unique paradigms or procedures. On the other hand, we should not be so specific in our requirements that labs and study types are immediately discounted. For example, requiring multitechnique imaging (e.g. EEG, fMRI) on every subject or certain electrode montages will preclude smaller labs and studies.
with younger children.

As a final word, no matter how stringent, any origin independent database will always require certain assumptions about the data. For example, we must assume that our fellow researchers are conscientious data collectors and do not alter or omit data. In addition we must assume a certain level of commonality in the characteristics of our described subject groupings and experimental settings (e.g. given that we want to generalize our results to young adults we must assume that data collected from undergraduates in California and New York are comparable). However, these assumptions are not specific to the creation of a database but are requirements for working within a scientific community.

In addition, the ability to use this database to make novel comparisons via "virtual" experiments will not preclude the need for actual experimentation and continued data collection. However, it can be used as a framework to guide the direction of research projects. Just as many experiments in physics are done as confirmation of mathematical predictions, the database can be used to predict experimental outcomes and increase the efficiency of neuroscience research.