

# Sistematização para massa de dados proveniente de sistemas externos de medição de grandezas relativas a energia primaria intermitente

## Design of the strategy for mass data from outdoor units of measurement of intermittent primary-energy quantities

Ananda Andrade NASCIMENTO [1](#); André Luiz Veiga GIMENES [2](#); Angélica Luana LINHARES [2](#); Antônio Celso ABREU Junior [3](#); Jonathas Luiz Oliveira BERNAL [2](#); Juliana de Oliveira Martins SPAGIARI [1](#); Luiz Claudio Ribeiro GALVÃO [2](#); Martim Debs GALVÃO [2](#); Miguel Edgar Molares UDAETA [2](#); Pascoal Henrique da Costa RIGOLIN [4](#); Paulo Helio KANAYAMA [5](#); Rafael Martinez ACEBRON [2](#); Raul Vaz de PAULA [6](#); Rodrigo Antônio CARNEIRO [2](#); Stefania Gomes RELVA [2](#); Vinícius Oliveira da SILVA [2](#);

Recebido: 28/08/2017 • Aprovado: 30/09/2017

### Conteúdo

- [1. Introdução](#)
- [2. Padronização de armazenamento de dados](#)
- [3. Procedimentos para armazenamento](#)
- [4. Tipos e formatos de armazenamento de dados](#)
- [5. Critérios de nomeação de arquivos](#)
- [6. Critérios de nomeação e disposição das informações dentro dos arquivos de massa de dados](#)
- [7. Procedimentos para disponibilização](#)
- [8. Conclusão](#)

[Agradecimentos](#)

[Referencias bibliográficas](#)

#### RESUMO:

Este artigo tem o objetivo de estabelecer um padrão de um conjunto de práticas e processos de uniformização de registro, armazenamento e disponibilização de dados de unidades de medição de grandezas relativas a energia primaria intermitente (solar e eólica). A metodologia definida envolve: (i) a análise exaustiva de procedimentos de armazenamento de dados de centros de referência internacionais e nacionais; (ii) ponderações e conhecimento sistêmico e consolidado de pesquisadores responsáveis pela utilização dos dados e

#### ABSTRACT:

This article aims to present a pattern of a set of practices and processes for the standardization of registry, storage and availability of data linked to primary intermittent energy (solar and wind energy). The defined methodology involves: (i) the thorough analysis of data storage assessment procedures from national and international reference centers; (ii) the considerations of researchers used to data management and responsible for the maintenance of solarimetric anemometric and meteorological stations maintenance

responsáveis pela manutenção de estações solarimétricas, anemométricas e meteorológicas no âmbito do GEPEA/EPUSP e (iii) a avaliação e análise específica dessas estações. Sendo apontados resultados tais como a idealização da apresentação de quatro massas de dados e a padronização de siglas e nomeação de arquivos e variáveis. Assim, conclui-se que essa formatação garante o acesso à informação de qualquer pessoa interessada na análise de dados solarimétricos, meteorológicos e anemométricos, independente da capacidade e experiência na análise desse tipo de variável, promovendo a facilitação de disponibilização de informações, tão necessárias ao desenvolvimento e aprimoramento do ambiente de investigação científica.

**Palavras-Chave:** Massa de dados, Registros de Medição, Energia Primária Intermitente, Padronização de Armazenamento de Dados

in the context of GEPEA/USP activities; (iii) the evaluation and specific analysis of these stations. As results we present an idealization of the organization of four mass data sets and a standardization of acronyms and naming of files and variables. Thus, it is concluded that this layout guarantees access to the information of anyone interested in the analysis of solarimetric, meteorological and anemometric data, regardless of the capacity and experience in the analysis of these types of variables, promoting a facility of availability information which is much-needed for the development and improvement of the scientific research environment.

**Keywords:** Mass data, Measurements registration, Primary intermittent energy, Data storage standardization

## 1. Introdução

O processo de pesquisa se caracteriza como o esforço cuidadoso para descoberta de novas informações ou relações e para a verificação e ampliação do conhecimento existente [1]. Quando se trata de pesquisa e investigação científica no campo dos sistemas elétricos e energéticos, a análise de desempenho e verificação do desenvolvimento é etapa crucial da conclusão do projeto. Parte essencial dessa verificação e do desenvolvimento de projetos é a disponibilização de dados confiáveis, ainda mais com a crescente utilização das energias renováveis intermitentes solar e eólica. Nesse sentido, verifica-se uma demanda cada vez maior do manuseio de dados por especialistas vinculados ao setor energético.

A partir das instalações dos meios de mensuração e da validação da confiabilidade desses meios, o armazenamento das informações disponíveis deve ser tratado com atenção, de modo a facilitar e possibilitar o desenvolvimento dos métodos científicos e análises de resultados.

Tratando-se especificamente projetos vinculados às novas fontes de energia e fontes renováveis e intermitentes, cujo desenvolvimento é recente no Brasil, carece-se da determinação de procedimentos de armazenamento dos dados vinculados aos recursos renováveis. Assim, esse estudo tem como principal objetivo apresentar um processo de determinação de armazenamento de massa de dados solarimétricos, anemométricos e meteorológicos, facilitando às empresas do setor, a implantação de um padrão de sistematização e disponibilização de dados que viabilize a completude de informações de massa de dados e a fácil compreensão dessas informações, de modo que o armazenamento dos históricos de medição sirva para e para além do meio científico.

A metodologia definida conta com a análise dos procedimentos de armazenamento de dados de centros de referência internacionais e nacionais e com os apontamento e sistematização do conhecimento de pesquisadores responsáveis pela utilização dos dados obtidos pelas referidas estações e responsáveis pela manutenção das mesmas.

## 2. Padronização de armazenamento de dados

A padronização do registro de medições é importante para que as interpretações dos dados possam ser corretas em relação: à medida a qual o dado se refere, às suas possíveis incertezas e ao horário de medição. O registro da hora, minuto e segundo de medição é um elemento que, se não for registrado em um padrão previamente acordado, pode gerar erros de interpretação e análise de dados.

A padronização do registro de dados, além de ser indispensável para a formação coerente de uma base regional, nacional ou mundial de dados, também é necessária no caso da utilização deste registro em softwares de análises de qualidade de dados.

Segundo a ANEEL [2], os registros de medição de uma estação solarimétrica devem conter, pelo menos, as seguintes informações:

Dia do início do intervalo (de 1 minuto) de medição

Horário do início do intervalo (de 1 minuto) de medição

Código de erro do intervalo, com "0" indicando medição sem erro

Pressão do ar [hPa]: média do intervalo de integração de 1 minutos

Temperatura do ar [°C]: média do intervalo de integração de 1 minuto

Umidade relativa do ar [%rel]: média do intervalo de integração de 1 minuto

Precipitação [mm]: média do intervalo de integração de 1 minuto

Velocidades do vento medidas pelo anemômetro [m/s]

Média do intervalo de integração de 1 minuto;

Máxima instantânea (1s) registrada no intervalo de integração de 1 minuto;

Mínima instantânea (1s) registrada no intervalo de integração de 1 minuto;

Desvio padrão no intervalo de integração de 1 minuto.

Direção do vento medida pelo Anemoscópio

Média do intervalo de integração de 1 minuto;

Desvio padrão no intervalo de integração de 1 minuto.

Irradiação global medida pelo piranômetro ou Piranômetro com Banda de Sombra [W/m<sup>2</sup>]

Média do intervalo de integração de 1 minuto;

Máxima instantânea (1s) registrada no intervalo de integração de 1 minuto;

Mínima instantânea (1s) registrada no intervalo de integração de 1 minuto;

Desvio padrão no intervalo de integração de 1 minuto.

Irradiação difusa medida pelo piranômetro sombreado ou o Piranômetro com Banda de Sombra [W/m<sup>2</sup>]

Média do intervalo de integração de 1 minuto;

Máxima instantânea (1s) registrada no intervalo de integração de 1 minuto;

Mínima instantânea (1s) registrada no intervalo de integração de 1 minuto;

Desvio padrão no intervalo de integração de 1 minuto.

Irradiação normal direta medida pelo pireliômetro ou o Piranômetro com Banda de Sombra [W/m<sup>2</sup>]

Média do intervalo de integração de 1 minuto;

Máxima instantânea (1s) registrada no intervalo de integração;

Mínima instantânea (1s) registrada no intervalo de integração;

Desvio padrão no intervalo de integração.

A ANEEL também define os padrões de nomenclatura dos arquivos de dados. Por exemplo, os dados da estação "000002" com período de dados entre 15/02/2012 e 16/02/2012 deve

possuir como nome do arquivo o seguinte código: "000002\_20120216\_20120215.txt". Os detalhes de siglas e padrões recomendados pela ANEEL podem ser verificados em [2].

Segundo a BSRN [3], o output final de cada variável radiométrica também deve consistir na média de medição de 1 minuto, o mínimo, máximo e o desvio padrão. A BSRN também indica que os dados de mínimo, máximo e desvio padrão não precisam ser enviados à base de dados centralizado, mas que devem ser registrados e manterem-se arquivados.

O principal objetivo da BSRN é fornecer medições de fluxos de radiação contínuas, de longos períodos, com o maior nível de qualidade, por meio das melhores práticas alcançáveis de medição, calibração e precisão. Para isso, o projeto vincula uma série de estações solarimétricas ao redor do mundo em uma base de dados. A manutenção de cada estação é de responsabilidade dos seus respectivos cientistas e instituições, de modo que, as estações vinculadas à BSRN não são de propriedade do projeto. Por consequência, a operação também não fica a encargo da BSRN. Portanto, essa configuração exige que seja estabelecido um processo padrão de registro de dados, que deve ser seguido pelos cientistas responsáveis por cada estação de medição.

Os arquivos de dados a serem enviados para a BSRN devem conter o número de identificação da estação, os dados do cientista responsável e a data do período de medição. Os dados devem ser enviados mensalmente. Os detalhes de siglas padrões utilizados são encontrados em [3]-[4].

O NREL define boas práticas de armazenamento de dados solarimétricos [5] e anemométricos [6].

Para definir a qualidade de dados de radiação o NREL desenvolveu um software denominado SERI QC que testa e sinaliza os dados de radiação que aparentam estar inconsistentes. O software foi desenvolvido para ser usado com dados de 1 minuto até dados de 60 minutos e tem sido utilizado por diferentes grupos ao redor do mundo, como pelo programa de medições de radiação atmosférica (ARM) do departamento norte-americano de energia. O software também tem sido utilizado em redes de medição na Coreia, Arábia Saudita e WMO.

Para a utilização do software os dados devem ser fornecidos em um determinado formato. O software requer informações iniciais de uma nova estação a ser registrada como: breve caracterização incluindo localização (cidade e estado); posição geográfica da estação (longitude e latitude em números decimais), fuso horário onde a estação está (o software aceita fuso horários decimais e define convenções de sinais para definição do fuso), com isso, um número de identificação da estação será gerado e esse número deve ser utilizado antes da inserção de novos documentos com registros de medição dessa estação. Para cada medida, o algoritmo do software requer o dado de irradiação solar em  $W/m^2$  e o horário da medição.

O manual para usuários do software [5] exemplifica o formato ideal de registro dos dados. O software desconsidera os valores decimais para datas e horários, mas considera duas casas decimais para os dados de radiação e para cada medição, deve-se indicar o horário quando a medição foi finalizada. Valores de medição perdidos devem ser indicados utilizando valores de dados acima de  $8.000 W/m^2$ . Por exemplo, uma medição que foi finalizada às 11 horas da manhã, do dia 12 de maio de 1979, deve ser registrada da seguinte maneira:

79 5 12 11 0 928.03 1102.47 9900.00

Os horários devem variar entre 00:01 à 24:00. E o software demanda dados de radiação difusa, direta e global para cada linha de dados. Nesse exemplo fornecido pelo manual, a medição foi finalizada no ano de 79, no mês 5, dia 12 às 11 horas e zero minutos e registrou uma irradiação global de  $928,03 W/m^2$ , a irradiação direta foi medida em  $1.102,47 W/m^2$  e o valor da componente difusa foi perdido/ não foi medido.

A WMO define boas práticas de medição e armazenamento de dados climatológicos [7], contudo, sem definir detalhes de siglas e padrões de uniformidade de registros.

### 3. Procedimentos para armazenamento

Assim a partir da familiaridade e pesquisa dos arquivos supracitados e das experiências de pesquisadores na análise de dados e manuseio dos mesmos, visando disponibilização de base de dados para interessados de diferentes áreas do conhecimento, definem-se os procedimentos de armazenamento desses dados. Estes procedimentos englobam: os tipos e formatos dos arquivos de armazenamento de dados (ii) as informações que devem estar presentes em cada arquivo, (iii) os critérios de nomeação de cada arquivo, (iv) a definição da ordem de disposição das informações, (v) os critérios de nomeação de cada variável e (vi) elementos de suporte ao acesso dos dados, como o meio de disponibilização dos mesmos e das demais informações que caracterizam cada estação de medição e disposição de diretórios.

---

### 4. Tipos e formatos de armazenamento de dados

Recomenda-se o armazenamento da mesma massa de dados em quatro diferentes modos, sendo estes: (i) dados físicos de base, (ii) dados brutos, (iii) dados originais formatados, (iv) dados originais amigáveis, basicamente o que difere os quatro tipos da mesma massa de dados é o nível de interferência aplicado à formatação da série histórica para o armazenamento. De modo que:

Os dados físicos de base são os dados de sinais elétricos (corrente ou tensão) enviados ao datalogger por cada sensor, ou seja, são os dados antes da aplicação dos fatores de conversão que relacionam o sinal elétrico a unidade de medida conveniente de cada sensor, essa massa de dados incorpora também os dados de controle. Os dados de controle são aqueles que apresentam o monitoramento do funcionamento da estação, e não a medição das variáveis externas. Tais como: nível da carga da bateria de alimentação da estação ao longo do tempo, alimentação da estação por corrente contínua ou alternada, performance dos ventiladores dos piranômetros ao longo do mês, dentre outros.

Os dados brutos são os dados já nas unidades de medidas convenientes de cada sensor ( $^{\circ}\text{C}$ , mm,  $\text{W}/\text{m}^2$ , etc). Os arquivos de dados brutos devem ser disponibilizados exatamente conforme foram obtidos pelo software vinculado ao datalogger de cada estação, sem nenhum tipo de formatação quanto as informações contidas dentro dos arquivos de massa de dados.

Os dados originais formatados correspondem aos dados brutos, porém com apenas uma linha de cabeçalho. Muitos arquivos de dados, quando obtidos dos softwares vinculados ao datalogger, originam um arquivo que, além da série histórica, também fornece um conjunto de linhas com informações técnicas acerca da estação. Essas linhas, quando exportadas à softwares de manuseio dos dados, podem desformatar a leitura tabelada da série histórica. Assim, nos arquivos de dados originais formatados são deletadas todas as linhas que não correspondem ao nome de cada variável medida ou linhas em branco entre o cabeçalho e os registros de medição. Nessa massa de dados as nomenclaturas da linha remanescente não devem ser alteradas.

Os dados originais amigáveis são aqueles que possuem uma formatação amigável das informações disponíveis na massa de dados, como: renomeação do cabeçalho, padronização da ordem de disposição de informações, exclusão de dados de controle dentre outros aspectos que são detalhados nas seções seguintes deste estudo.

Recomenda-se que os arquivos das massas de dados 3 e 4 (dados originais formatados e dados originais amigáveis) sejam em formato .txt já que a extensão é adaptável à utilização em diversos softwares. Já as massas de dados 1 e 2 (dados físicos de base e dados brutos) devem ser disponibilizados no mesmo formato obtido através do software vinculado à cada estação de medição, normalmente encontrados em .csv, .xls ou .txt.

Cada estação de medição é constituída de um conjunto de sensores que medem diferentes

variáveis. Portanto recomenda-se que, as massas de dados 1 e 2 sejam disponibilizadas em arquivos mensais de cada estação; a massa de dados 3 seja disponibilizada em arquivos diários de cada estação e, ao final de um mês completo de medições forneça-se também um arquivo .txt completo, ou seja, com as medições do mês inteiro.

A massa de dados 4 deve ser disponibilizada em arquivos diários separados para cada sensor de cada estação e ao final de um mês completo de medições deve-se fornecer também um arquivo .txt completo: do mês inteiro de medições com todas as variáveis de medição de cada estação de medição.

As medições a serem registradas devem prioritariamente ser fornecidas de minuto a minuto.

Para a formatação da massa de dados 4 é importante observar que as informações contidas em cada arquivo de dados devem obedecer ao seguinte critério: caso o sensor registre medições de segundo em segundo, o arquivo de massa de dados deve registrar a média de cada minuto, seguido do desvio padrão deste período e dos respectivos máximo e mínimo. Caso o datalogger promova também o registro de contador, ou seja, apresente o valor de quantas medidas foram registradas naquele minuto, essa informação também deve ser fornecida no arquivo .txt daquele sensor. Além disso, o arquivo deve conter a data completa de cada medição: ano, mês, dia, hora, minuto, segundo e dia juliano, além do formato padrão de datas fornecido pelo datalogger.

Tanto para arquivos de dados diários quanto para o arquivo do conjunto mensal de dados das massas de dados 3 e 4 é importante que haja apenas uma linha de cabeçalho, na qual deve conter apenas o nome das variáveis apresentadas. Cada momento de medição, com os respectivos valores medidos naquele momento devem constar em linhas subsequentes ao cabeçalho.

---

## 5. Critérios de nomeação de arquivos

O nome da estação deve ser definido em 5 letras em caixa alta, de modo que as duas primeiras definam o tipo da estação conforme o seguinte padrão:

- ES: Estação Solarimétrica
- EA: Estação Anemométrica
- EM: Estação Meteorológica

As três letras seguintes devem significar o nome que o definido para referenciar a estação em questão.

A nomeação de todos os arquivos, de todas as massas de dados (1,2,3 ou 4) deve sempre ser iniciada com a sigla da estação que originou os dados. As demais informações da nomeação dos arquivos são detalhadas a seguir. Cada informação do nome deve ser diferenciada da próxima informação por meio da utilização do caractere underscore (\_).

Os arquivos mensais das massas de dados 1 e 2 devem possuir as seguintes informações no nome do arquivo, na ordem apresentada a seguir:

1. Nome da estação
2. Data
3. Período de integração dos dados
4. Tipo de massa de dados

Os arquivos diários da massa de dados 3 deve possuir as seguintes informações, na ordem apresentada a seguir:

1. Nome da estação
2. Latitude da estação
3. Longitude da estação

4. Altitude da estação
5. Data (ano, mês, dia)
6. Período de integração dos dados
7. Tipo de massa de dados

Enquanto que os arquivos mensais devem possuir as informações:

1. Nome da estação
2. Latitude da estação
3. Longitude da estação
4. Altitude da estação
5. Data (ano, mês)
6. Período de integração dos dados
7. Tipo de massa de dados

Os arquivos diários da massa de dados 4 deve possuir as seguintes informações, na ordem apresentada a seguir:

1. Nome da estação
2. Latitude da estação
3. Longitude da estação
4. Altitude da estação
5. Nome do sensor
6. Nome da variável de medição
7. Data (ano, mês, dia)
8. Período de integração dos dados
9. Tipo de massa de dados

Já os arquivos mensais devem apresentar as seguintes informações:

1. Nome da estação
2. Latitude da estação
3. Longitude da estação
4. Altitude da estação
5. Nome da variável de medição
6. Data (ano, mês, dia)
7. Período de integração dos dados
8. Tipo de massa de dados

A latitude deve ser definida com uma letra em caixa alta (N ou S), dois números inteiros e três casas decimais. A separação entre inteiros e decimais deve ser feita por hífen (-), já que pontos (.) podem causar erros de leituras em alguns softwares e sistemas computacionais.

Assim como a latitude, a longitude deve ser definida com uma letra em caixa alta (L ou O), dois números inteiros e três casas decimais. A separação entre inteiros e decimais deve ser feita por hífen (-), já a altitude deve ser definida com quatro números inteiros.

Para a massa de dados 4, que define a disponibilidade de arquivos diários de dados por sensor, é necessária a sigla do sensor e da variável medida no nome do arquivo, assim recomenda-se que o nome do sensor seja definido em duas letras com a primeira em caixa alta. Caso seja necessária a definição da altura de um sensor para diferenciação de sensores iguais em uma mesma estação (como no caso de anemômetros que comumente são instalados em diversas

alturas em uma mesma estação de medição), o valor da altura do equipamento deve ser colocado imediatamente após a sigla do nome do equipamento sem a utilização do caractere underscore (\_). O valor da altura deve ser sempre registrado com três números inteiros. Assim, se altura do equipamento for da ordem de dezenas, o primeiro valor dos três números inteiros deve ser zero (0).

A mesma lógica se aplica para a altitude. Caso a altitude da estação seja da ordem de dezenas, a informação deve ser iniciada com dois zeros (00), caso seja da ordem de centenas, deve ser iniciada com um zero (0). A Tabela I apresenta uma sequência sugerida de siglas para os equipamentos definidos para estações solarimétricas, anemométricas e climatológicas.

**Tabela I**  
Padrão de Siglas de Equipamentos.

<b>Equipamento</b>	<b>Sigla</b>
Anemômetro	Am
Anemoscópio	Ac
Barômetro	Br
Piranômetro	Pi
Pireliômetro	Pr
Pluviômetro	Pl
Termo-Higrômetro	Th

Fonte: Elaboração própria

O nome da variável de medição deve ser definido em duas letras com a primeira em caixa alta. Essa informação no nome dos arquivos diários da massa de dados 4 é considerada visto que alguns sensores podem medir mais do que uma variável, assim como algumas variáveis podem ser obtidas por sensores diferentes. Tabela II apresenta uma sequência sugerida de siglas para as variáveis comumente medidas em estações solarimétricas, anemométricas e climatológicas.

**Tabela II**  
Padrão de Siglas de Variáveis.

<b>Equipamento</b>	<b>Sigla</b>
Velocidade do Vento	Vv
Direção do Vento	Dv
Pressão	Pr
Irradiancia Global Horizontal (IGH)	GI
Irradiancia Difusa Horizontal (IDH)	Df



Irradiância Direta Normal (IDN)	Dr
Precipitação	Pp
Temperatura	Tp

Fonte: Elaboração própria

O nome do arquivo de dados completos mensais das massas de dados 3 e 4 deve desconsiderar o dia, o nome do equipamento e o nome da variável (já que todas as variáveis devem estar incluídas no mesmo arquivo).

Todos os arquivos de massas de dados devem possuir a informação da data do período de medição na sua nomeação. A data informada no nome dos arquivos de dados diários deve seguir o formato AAAAMMDD, enquanto que a data informada no nome dos arquivos de dados mensais deve seguir o formato AAAAMM.

Outro aspecto temporal a ser definido no nome dos arquivos de massas de dados é o período de integração dos dados. O período de integração de dados corresponde à unidade de tempo na qual os dados são apresentados. No caso ideal, consideram-se arquivos com dados brutos de minuto a minuto. Contudo, é possível a obtenção de registros de segundo em segundo, dentre outros períodos. Assim, determina-se que o registro deve ser prioritariamente de minuto a minuto. Logo, a informação que deve ser inserida no nome dos arquivos de todas as massas de dados é uma letra em caixa alta, sendo: "M" para dados integrados ou registrados de minuto a minuto e "S" para dados registrados de segundo a segundo.

A última informação a ser adicionada a todos os arquivos de histórico de medição é a qual massa e dados aquele arquivo faz parte. Essa informação deve ser colocada através da adição de uma letra em caixa alta ao final do arquivo. Assim, define-se conforme a Tabela III a letra referente a cada tipo.

**Tabela III**

Letras referentes a cada massa de dados.

<b>Massa de Dados</b>	<b>Letra para nomeação do arquivo</b>
1- Dados físicos de base	F
2- Dados brutos	B
3- Dados originais formatados	O
4- Dados originais amigáveis	A

Fonte: Elaboração própria

Considerando uma estação anemométrica (EA) denominada de Usina Porto Primavera (UPP), cuja latitude é 22,48° Sul, a longitude é de 52,95° Oeste e altitude é de 284 metros, os arquivos de dados mensais de março de 2016 e do anemômetro instalado à 80 metros, de altura considerando dados diários do dia 07, deve seguir os exemplos apresentados na Tabela IV.

**Tabela IV**

<b>Dados</b>	<b>Dados diários por equipamento</b>	<b>Dados diários completos</b>	<b>Dados Mensais Completos</b>
1	Não disponível	Não disponível	EAUPP_201603_M_F
2	Não disponível	Não disponível	EAUPP_201603_M_B
3	Não disponível	EAUPP_S22-482_O52 954_0284_20160307_M_O	EAUPP_S22-482_ O52954_0284_201603 _M_O
4	EAUPP_S22-482_O52-954_ 0284_Am080_Vv_20160307 _M_A	Não disponível	EAUPP_S22-482_ O52954_0284_201603 _M_A

Fonte: Elaboração própria

## 6. Critérios de nomeação e disposição das informações dentro dos arquivos de massa de dados 4.

A massa de dados 4 (dados originais formatados) é a única definida com alterações internas nos arquivos de dados. Nos arquivos de massa de dados 3 apenas devem ser deletadas linhas que não correspondam à linha original de nome das variáveis, contudo, essa linha remanescente não deve sofrer alterações. Assim, essa seção apresenta as alterações que devem ser realizadas dentro dos arquivos da massa de dados 4.

É importante ressaltar que a massa de dados 4 condiz estritamente aos valores das variáveis medidas pela estação, de modo que qualquer dado de controle ou físico bruto deve ser excluído desses arquivos de massas de dados.

Para que a compreensão das informações apresentadas seja facilitada, os arquivos .txt da massa de dados 4 devem possuir todos a mesma estrutura interna. Dados registrados de um em um minuto podem geralmente ser originados (i) da média da integração de 60 medições (por segundo), (ii) de apenas um registro por minuto ou (iii) da integração de 60 medições (por segundo). Em qualquer um dos casos, a estrutura de dados deve conter, antes dos registros de medições, as informações de datas e horários.

O datalogger normalmente registra a data em um único vetor, no formato: AAAAMMDD-HH:MM:SS. Essa estrutura deve ser padronizada dividindo as informações em vetores diferentes, com nomes padronizados, conforme o cabeçalho apresentado na Figura 1.

**Figura 1**  
Cabeçalho de disposição de datas.

<b>Data</b>	<b>Ano</b>	<b>Mes</b>	<b>Dia</b>	<b>Hora</b>	<b>Minuto</b>	<b>Segundo</b>	<b>Dia_J</b>
-------------	------------	------------	------------	-------------	---------------	----------------	--------------

Fonte: Elaboração própria

A variável Data representa o formato original de data fornecido pelo datalogger, que não deve ser apagado ou editado.

O ano deve ser registrado com 4 números, o mês e o dia com dois números. A informação

'mês' deve ser registrada no arquivo .txt sem acento, como na Figura 1. Todos os nomes devem ser inicializados com uma letra em caixa alta.

As horas devem variar entre 00 e 23 e os minutos devem variar entre 00 e 59, assim como os segundos. De modo que o primeiro horário do dia seja 00:00:00 e o último seja 23:59:59 no caso de arquivos com dados registrados por segundo e 23:59:00 no caso de arquivos com dados registrados por minuto. Ou seja, nas tabelas com dados por minuto os segundos devem estar sempre zerados, a não ser que o datalogger tenha registrado um valor de segundo diferente, nesse caso, esse valor deve ser colocado na variável 'Segundo' já que essa informação pode representar um erro de registro naquele momento.

A variável Dia\_J representa o dia juliano e varia de 1 a 365. Em caso de ano bissexto o a variável varia de 1 a 366.

Após a definição dessas informações, nessa ordem, procede-se às informações dos dados medidos. Em caso de dados que são coletados por minuto originados da média da integração de 60 medições (por segundo), o datalogger fornece a média do minuto (Avg), os valores máximo (Max) e mínimo (Min) das medições ao longo desse minuto e o desvio padrão (Std). Nesse caso, o registro da variável deve iniciar com a sigla da variável, como já definido na Tabela II, seguido do caractere underscore (\_) e da sigla do tipo de medida.

Para variáveis cujo o registro é feito pontualmente (Int) uma vez por minuto, sem que aquela medição represente a média do acúmulo ou o próprio acúmulo ao longo do minuto, como, por exemplo, pode ocorrer com a temperatura ambiente, a notação a ser utilizada deve ser a apresentada a sigla da variável, como já definido na Tabela II, seguido do caractere underscore (\_) e da sigla do tipo de medida (Int). No caso de variáveis como a precipitação, por exemplo, que se costuma apresentar o total acumulado (Sum) em um minuto, a notação a ser utilizada deve ser a apresentada a sigla da variável, como já definido na Tabela II, seguido do caractere underscore (\_) e da sigla do tipo de medida (Sum).

Algumas estações de medição também fornecem um dado extra, definido como Contador. O Contador indica quantos registros foram realizados para que se alcançasse o valor de medição apresentado. Tomando como exemplo a irradiância global: esse tipo de variável pode ser obtido por segundo, portanto 60 medições devem ser utilizadas para o cálculo dos valores de média, máximo, mínimo e desvio padrão por minuto. Nesse caso o contador de cada minuto deve ser igual à 60. Assumindo o exemplo da temperatura, que pode ter medições pontuais por minuto: nesse caso o contador deve assumir o valor igual à 1 para cada minuto. Em caso do datalogger disponibilizar essa informação, ela deve ser a última a ser apresentada no conjunto de informações daquele sensor/variável.

Assim, a Tabela V apresenta sistematicamente a sigla para cada tipo de valor apresentado. Caso algum novo tipo de valor ocorra, uma nova sigla precisa ser criada e registrada em uma nova versão do documento de especificação do registro de dados. Qualquer nova sigla deve obedecer ao padrão de possuir 3 letras com a primeira em caixa alta. As siglas já existentes não devem ser alteradas já que isso acarretaria em prejuízo à padronização de dados existentes e futuros.

**Tabela V**

Padrão de sigla para cada tipo de valor de variáveis medidas.

<b>Tipo de Valor</b>	<b>Sigla</b>
Média de 60 medições realizadas em 1 min.	Avg
Valor máximo de 60 medições realizadas em 1 min.	Max
Valor mínimo de 60 medições realizadas em 1 min.	Min

Desvio padrão de 60 medições realizadas em 1 min.	Std
Integral dos valores obtidos em 1 min.	Sum
Registro pontual realizada 1 vez por min.	Int
Contador	Cnt

Fonte: Elaboração própria.

Os vetores de informações devem estar dispostos no arquivo .txt necessariamente na ordem apresentada nessa seção. Cada vetor deve ser separado por “,”. É importante observar que esse tipo de alteração, quando realizada diretamente no arquivo .txt é passível de alguns erros e deformações, portanto sugere-se que esse tipo de alteração interna seja realizado com a utilização de software apropriado para tratamento de massas de dados, são exemplos a utilização do Microsoft Excel e o R, dentre outros.

Valores inexistentes de medições devem estar registrados como “NA”.

No caso do arquivo de dados mensais, é importante que a sequência de apresentação das diferentes variáveis medidas (com os seus respectivos tipos de valor) seja a mesma em todos os arquivos de todas as estações. Assim define-se a ordem de disposição das variáveis, de modo que a primeira variável a ser apresentada é aquela alocada logo após aos vetores de data apresentados na Figura 1.

## **1. Estação Meteorológica**

- a. Velocidade do Vento
- b. Direção do Vento
- c. Temperatura
- d. Pressão atmosférica
- e. Precipitação
- f. Umidade Relativa

## **2. Estação Anemométrica**

- a. Velocidade do Vento da menor altura de instalação do medidor
- b. Direção do Vento da menor altura de instalação do medidor
- c. Velocidade do Vento da altura intermediária de instalação do medidor
- d. Direção do Vento da altura intermediária de instalação do medidor
- e. Velocidade do Vento da maior altura de instalação do medidor
- f. Direção do Vento da maior altura de instalação do medidor
- g. Temperatura
- h. Pressão atmosférica
- i. Precipitação
- j. Umidade Relativa

No caso de a estação possuir mais de 3 conjuntos de Anemômetro/Anemoscópio, procede-se com a mesma lógica. Essa lógica também deve ser aplicada para estações solarimétricas e meteorológicas que eventualmente tenham mais do que um conjunto de anemômetro/anemoscópio.

### 3. Estação Solarimétrica

- a. Irradiância Global Horizontal
- b. Irradiância Difusa Horizontal
- c. Irradiância Direta Normal
- d. Velocidade do Vento
- e. Direção do Vento
- f. Temperatura
- g. Pressão atmosférica
- h. Precipitação
- i. Umidade Relativa

No caso de a estação possuir outros medidores de radiação, esses devem ser apresentados logo após os dados de Irradiância Direta Normal e antes dos dados de Velocidade do Vento.

Tratando-se de estação solarimétrica é importante que os dados de irradiância ou aqueles vinculados ao brilho solar sejam os primeiros apresentados, a mesma lógica se aplica à estação anemométrica no caso de dados de ventos. As demais variáveis meteorológicas tiveram a ordem padronizada para todas as estações, evitando diferenças entre arquivos de estações diferentes. Portanto, no caso do acréscimo de novas variáveis, a lógica estabelecida é: encaixar na ordem conforme a relevância do dado, levando em consideração qual o principal fim da estação. Contudo, a ordem definida para a apresentação dos dados de velocidade e direção de vento, bem como de irradiâncias global, direta e difusa não devem deixar de serem apresentadas prioritariamente para estações anemométricas e solarimétricas respectivamente.

---

## 7. Procedimentos para disponibilização

Os dados devem ser disponibilizados preferivelmente em servidor com a criação de link para acesso externo, possibilitando e favorecendo o acesso à informação. Cada estação deve possuir um diretório próprio, e dentro deste, outros diretórios que devem armazenar os dados sistematizados por data e por variável, além das informações características da estação, tais como: listagem dos equipamentos instalados, incluindo marca e modelo; caracterização da localização da estação; arquivo com as unidades de medida de cada variável apresentada, bem como a definição das siglas dos equipamentos e variáveis.

Dentro dos diretórios de massas de dados deve-se armazenar, além dos dados, um arquivo .txt que explique sucintamente qual a configuração adotada para cada uma das quatro massas de dados, conforme as definições já apresentadas nesse estudo.

Assim, sugere-se a sistematização dos diretórios da seguinte maneira:

**Diretório de Estações de Medição:** esse é diretório inicial que deve concentrar as informações de todas as estações de medição das quais o projeto possui acesso. Esse diretório deve ser nomeado como: "Estações de Medição"

**Arquivo.doc-Armazenamento de dados:** inserido no diretório Estações de Medição e corresponde ao arquivo de informações contidas nesse estudo. Esse arquivo deve ser nomeado como: "Procedimento de Armazenamento de Dados\_ 'AAAA'". A cada atualização deve ser salva uma nova versão, indicando, no nome do arquivo, o ano da atualização. Em caso de mais de uma atualização no mesmo ano, deve-se acrescentar, no nome do arquivo, uma letra minúscula do alfabeto latino, iniciando-se em "b", para a segunda versão do ano.

**Diretório de cada estação:** cada estação da qual há acesso aos dados deve ter o seu próprio diretório dentro do diretório "Estações de Medição". Nele deverão ser incluídas de maneira sistemática (conforme itens a seguir) todas as informações referentes à

estação. Esse diretório deve ser nomeado seguindo o padrão: “‘Nome da estação’ (‘Sigla da estação’) “. A sigla deve ser definida conforme o padrão apresentado na seção B deste estudo.

**Diretório de dados:** nesse diretório deve haver todas as informações referentes ao histórico de medições da estação. Ele deve ser nomeado como: “Dados” e deve estar inserido no diretório de cada estação de medição. Dentro do diretório de dados devem haver quatro diretórios principais referentes a cada uma das quatro massas de dados.

**Diretório de dados físicos de base:** deve estar inserido no diretório de dados de cada estação. Nesse diretório deve haver toda a massa de dados 1 da estação de medição. Ele deve ser nomeado como: “1 Dados Físicos de Base”.

**Arquivo .txt com a descrição do que são os dados físicos de base:** esse arquivo deve estar inserido no diretório de dados físicos de base e deve possuir um texto explicando sucintamente o que são os dados físicos de base e o nível de interferência realizado no arquivo antes da sua disponibilização (de preferência nenhuma outra interferência além da renomeação do arquivo). A nomeação do arquivo deve ser: “Caracterizacao\_Dados\_Físicos”.

**Diretórios de dados de cada ano:** dentro do diretório de dados físicos de base deve haver um diretório para cada ano de medição da estação. A nomeação desse diretório deve seguir o padrão: “‘Sigla da esta-ção’\_‘AAAA’\_F”.

**Arquivos de todos os dados físicos de base de cada mês do ano.** Esses arquivos devem estar inseridos no diretório do ano correspondente e devem ser nomeados conforme as especificações tratadas na seção B desse estudo.

**Diretório de dados brutos:** deve estar inserido no diretório de dados de cada estação. Nesse diretório deve haver toda a massa de dados 2 da estação de medição. Ele deve ser nomeado como: “2 Dados Brutos”

**Arquivo .txt com a descrição do que são os dados brutos:** esse arquivo deve estar inserido no diretório de dados brutos e deve possuir um texto explicando sucintamente o que são os dados brutos e o nível de interferência realizado no arquivo antes da sua disponibilização (de preferência nenhuma outra interferência além da renomeação do arquivo). A nomeação desse arquivo deve ser: “Caracterizacao\_Dados\_Brutos”.

**Diretórios de dados de cada ano:** dentro do diretório de dados brutos da estação deve haver um diretório para cada ano de medição da estação. A nomeação desse diretório deve seguir o padrão: “‘Sigla da esta-ção’\_‘AAAA’\_B”.

**Arquivos de todos os dados brutos de cada mês do ano:** esses arquivos devem estar inseridos no diretório do ano correspondente e devem ser nomeados conforme as especificações tratadas na seção B desse estudo.

**Diretório de dados originais formatados:** deve estar inserido no diretório de dados de cada estação. Nesse diretório deve haver toda a massa de dados 3 da estação de medição. Ele deve ser nomeado como: “3 Dados Originais Formatados”

**Arquivo .txt com a descrição do que são os dados originais formatados:** esse arquivo deve estar inserido no diretório de dados originais formatados e deve possuir um texto explicando sucintamente o que são os originais formatados e o nível de interferência realizado no arquivo antes da sua disponibilização (de preferência nenhuma outra interferência além da renomeação do arquivo e da retirada de linhas de textos do arquivo que não correspondam ao cabeçalho de uma linha da estrutura de dados). A nomeação do arquivo deve ser: “Caracterizacao\_Dados\_Originais”.

**Diretórios de dados de cada ano:** dentro do diretório de dados originais formatados deve haver um diretório para cada ano de medição da estação. A nomeação desse diretório deve seguir o padrão: “‘Sigla da esta-ção’\_‘AAAA’\_O”.

**Diretórios de dados de cada mês do ano:** dentro do diretório de cada ano dos dados originais formatados deve haver uma quantidade de diretórios igual à quantidade de meses do ano que a estação registrou medições. A nomeação desses diretórios mensais deve seguir o padrão: “‘Sigla da estação’\_‘AAAAMM’\_O”.

**Arquivos .txt de todos os dados originais formatados de cada dia do mês:** esses arquivos devem estar inseridos no diretório do mês correspondente e devem ser nomeados conforme as especificações tratadas na seção B desse estudo.

**Arquivo .txt de todos os dados os dados originais formatados do mês:** esse arquivo deve estar inserido no diretório do mês correspondente e deve ser nomeado conforme as especificações tratadas na seção B desse estudo.

**Diretório de dados originais amigáveis:** deve estar inserido no diretório de dados de cada estação. Nesse diretório deve haver toda a massa de dados 4 da estação de medição. Ele deve ser nomeado como: “4 Dados Originais Amigáveis”

**Arquivo .txt com a descrição do que são os dados originais amigáveis:** esse arquivo estar inserido no diretório de dados originais amigáveis e deve possuir um texto explicando sucintamente o que são os dados originais amigáveis e o nível de interferência realizado no arquivo antes da sua disponibilização. A nomeação deve arquivo deve ser: “Caracteriza-cao\_Dados\_Amigáveis”.

**Arquivo .tx. de unidade de medidas e siglas:** deve ser gerado um arquivo .txt com vetores separados por apenas “,”, sem a utilização de nenhum espaço ou quaisquer outros diacríticos. O primeiro vetor deve conter a lista de variáveis de medição da estação, o segundo vetor deve conter as respectivas siglas de cada variável conforme definido nas seções B e C deste estudo. O terceiro vetor deve conter as respectivas unidades de medição nas quais as medições estão sendo disponibilizadas. Caso a estação forneça o valor “Contador” é importante que esse arquivo .txt também indique que a unidade de medição desse valor é unitária adimensional. Esse arquivo deve ser nomeado como “‘Sigla da Esta-ção’\_Unidades\_de\_Medidas\_e\_Siglas”. Esse arquivo deve conter 1 linha de cabeçalho, os 3 vetores devem ser nomeados respectivamente como: “Variavel”, “Sigla”, “Unidade\_de\_Medicao”.

**Diretórios de dados de cada ano:** dentro do diretório de dados originais amigáveis deve haver um diretório para cada ano de medição da estação. A nomeação desse diretório deve seguir o padrão: “‘Sigla da estação’\_‘AAAA’\_A”.

**Diretórios de dados de cada mês do ano:** dentro do diretório de cada ano dos dados originais amigáveis deve haver uma quantidade de diretórios igual à quantidade de meses do ano que a estação registrou medições. A nomeação desses diretórios mensais deve seguir o padrão: “‘Sigla da estação’\_‘AAAAMM’\_A”.

**Arquivo .txt de todos os dados os dados originais formatados do mês:** esse arquivo deve estar inserido no diretório do mês correspondente e deve ser nomeado conforme as especificações tratadas na seção B desse estudo.

**Diretórios de dados de cada dia do mês:** dentro do diretório mensal deve haver uma quantidade de diretórios igual à quantidade de dias do mês que a estação registrou medições. A nomeação desses diretórios diários deve seguir o padrão: “‘Sigla da estação’\_‘AAAAMMDD’\_A”.

**Arquivos .txt de dados originais amigáveis de cada variável do dia:** esses arquivos devem estar inseridos no diretório do dia correspondente e devem ser nomeados conforme as especificações tratadas na seção B desse estudo.

**Diretório de arquivos auxiliares:** este diretório deve estar inserido no diretório da estação de medição (junto com o diretório de dados) e deve conter as especificações técnicas da estação de medição. Ele deve ser nomeado como: “Especificações da Estação”.

**Arquivo.txt de especificação de equipamentos e da localização da estação:** deve estar inserido no diretório de arquivos auxiliares e deve conter as informações (sem padronização de vetores): (i) localização da estação-latitude, longitude, altitude, (ii) lista de equipamentos instalados e seus respectivos modelos, marcas e siglas utilizadas nos arquivos de dados e (iii) se possível, breve descrição dos elementos ao redor do local aonde a estação está instalada (ex. presença de árvores, prédios, altos níveis de poluição e etc). Esse arquivo deve ser nomeado como: "Sigla da estação'\_Caracterizacao"

**Demais diretórios:** caso o projeto em questão possua acesso aos manuais dos equipamentos e demais documentos dessa ordem, deverá ser criado um novo diretório dentro do diretório "Especificações da Estação", para o armazenamento desses documentos.

---

## 8. Conclusão

Esse estudo buscou sistematizar a apresentação e nomeação tanto dos dados, quanto da disponibilização dos dados de energia primária intermitente provenientes de estações solarimétricas, anemométricas e meteorológicas. Deste modo, busca-se a implementação desse processo para que o fluxo de informações seja perene e eficaz.

Os processos apresentados aqui foram construídos com base na experiência de pesquisadores no manuseio dos dados visando projetos energéticos e na análise de publicações relevantes ao tema, de modo que tais definições englobam um conjunto de sugestões de processos de registro, padronização, armazenagem e disponibilização de dados que podem ser alterados conforme a necessidade de cada projeto ou área de interesse. Recomenda-se, contudo, que quaisquer alterações nos processos definidos nesse estudo, se aplicados, devem ser documentadas, apresentando o novo padrão e indicando o padrão anterior, para que também não haja perdas do fluxo de informação desse processo. É importante ressaltar que, para qualquer novo padrão de nomeação de arquivos e variáveis, deve-se sempre evitar a utilização de caracteres como: ponto (.), acentos (^) e quaisquer outros diacríticos já que esses sinais podem prejudicar a leitura dos dados em softwares específicos de análise e tratamento dos mesmos.

A massa de dados 4, configurada para a apresentação dos dados mais amigáveis possíveis, ainda que formatada, não possui quaisquer procedimentos de análise de qualidade de dados aplicados. Assim, essa formatação também não garante a completude da série histórica de cada arquivo disponível, de modo que pode ocorrer repetição de linhas de dados, ausência de registros de temporais ou desordem dos mesmos.

Assim, conclui-se que a padronização e definição de quatro massas de dados garante o acesso à informação de qualquer pessoa interessada na análise desses, independente da capacidade e experiência no manuseio desse tipo de variável, assim promove-se a facilitação de disponibilização de informações, tão necessárias ao desenvolvimento e aprimoramento da pesquisa, do conhecimento e de projetos na área da energia.

## Agradecimentos

À CESP por ser financiador dos P&Ds ANEEL PD-0061-0043/2014, ANEEL PE-0061-0037/2012 e ANEEL PE-0061-0034/2011, que possibilitaram o desenvolvimento deste artigo. E também a equipe de pesquisadores do GEPEA/EPUSP e colaboradores que participaram direta e indiretamente na realização destes P&Ds ANEEL.

---

## Referencias bibliográficas

[1] A. S. Godoy, "Introdução à Pesquisa Qualitativa e suas Possibilidades", Revista de Administração de Empresas, vol. 35, n.2, pp.57-63, Mar/Abr 1995.



[2] ANEEL. Superintendência de Pesquisa e Desenvolvimento e Eficiência Energética – SPE. “Estação solarimétrica e equipamento de medição associados à Chamada de Projeto de P&D Estratégico Nº 013/2011.FI”. 02 do Ofício Circular nº 0004/2012-SPE/ANEEL, de 16/08/002.

[3] L. J. B. McArthur, “Baseline Surface Radiation Network (BSRN): operations manual, version 2.1.” Geneva: WCRP-121; WMO/TD-No. 1274, 2005.

[4] WMO [World Meteorological Organization]. World Climate Research Program Report: Baseline Surface Radiation Network (BSRN), update of the technical plan for BSRN Data Management. Geneva: Chairperson, Publications Board, 2013.

[5] E. Maxwell, S. Wilcox, M. Rymes, “Users manual for SERI QC software, assessing the quality of solar radiation data.” Report No. NREL-TP-463-5608. 1617 Cole Boulevard, Golden, Colorado, National Renewable Energy Laboratory, 1993.

[6] NREL [National Renewable Energy Laboratory], Wind Source Assessment Handbook - Fundamentals for Conducting a Successful Monitoring Program. Albany, 1997.

[7] WMO [World Meteorological Organization] Guide to Meteorological Instruments and Methods of Observation, Ed. de 2008, atualizada em 2010. Geneva: Chairperson, Publications Board, 2012.

- 
1. Instituto Federal de Educação (IFSP), ciência e tecnologia de São Paulo, Brasil
  2. Grupo de Energia do Departamento de Engenharia de Energia e Automação Elétricas da Escola Politécnica da Universidade de São Paulo (GEPEA/EPUSP), São Paulo, Brasil
  3. Secretaria de Energia e Mineração do Estado de São Paulo, São Paulo, Brasil.
  4. Ryerson University, Toronto, Canadá
  5. Centro Estadual de Educação Tecnológica Paula Souza, Faculdade de Tecnologia de Itaquera (FATEC), São Paulo, Brasil
  6. Universidade Estadual Paulista Júlio de Mesquita Filho – Campus Ilha Solteira/SP (UNESP), São Paulo, Brasil
- \*. Autor de correspondência. Email: [vinicius.oliveira.silva@usp.br](mailto:vinicius.oliveira.silva@usp.br)\*
- 

Revista ESPACIOS. ISSN 0798 1015  
Vol. 39 (Nº 09) Año 2018

[Índice]

[No caso de você encontrar quaisquer erros neste site, por favor envie e-mail para [webmaster](#)]

©2018. revistaESPACIOS.com • Derechos Reservados