|  |
| --- |
| Brandt Meio Ambiente |
| Padronização Banco de Dados |
| [Subtítulo do documento] |

|  |
| --- |
| Francisco de Paula Ribeiro  11/12/2021 |

Sumário

[1 – FORMATAÇÃO BÁSICA 4](#_Toc90226010)

[2 – NOME DO ARQUIVO 4](#_Toc90226011)

[3 – PLANILHA ÚNICA 4](#_Toc90226012)

[4 – IDENTIFICADOR 5](#_Toc90226013)

[5 – VARIÁVEIS POR COLUNAS 6](#_Toc90226014)

[6 – VARIÁVEIS POR LINHAS 7](#_Toc90226015)

[7 – CODIFICAÇÃO 7](#_Toc90226016)

[8 – COORDENADAS 9](#_Toc90226017)

[9 – DATA E HORA 9](#_Toc90226018)

[10 – DADOS ACUMULADOS 9](#_Toc90226019)

APRESENTAÇÃO

O presente documento foi elaborado com o objetivo de orientar a padronizar o banco de dados da Brandt Meio Ambiente, procurando evitar a ocorrência de vícios nas bases de dados.

Inicialmente é realizada uma breve descrição sobre formatação básica. Por fim, são apresentados alguns padrões de formatação específicos.

Com este manual, os colaboradores passam a dispor de uma fonte permanente de consulta de orientações sobre como padronizar, como formatar as planilhas ao armazenar os dados em arquivo Excel.

Trata-se de uma contribuição inaugural, que obviamente poderá demandar futuras ampliações e reedições visando ao aprimoramento do presente manual, mas que objetiva, desde logo, a suprir uma necessidade sentida pela organização. Adotar uma padronização dos dados é parte de um esforço em contribuir para uma administração mais organizada e eficaz.

# 1 – FORMATAÇÃO BÁSICA

Os bancos de dados armazenados em planilha Excel devem seguir a seguinte formatação básica:

**Fonte:** Arial

**Tamanho da fonte:** 12 pt

**Altura da linha:** 15

**Alinhamento do texto:** centralizado

# 2 – NOME DO ARQUIVO

Para nomearmos um banco de dados em arquivo Excel, adotaremos o seguinte padrão: **NÚMERO DO RELATÓRIO CONSOLIDADO CÓDIGO DO PROGRAMA/PROJETO NOME DA PLANILHA (DATA DA ÚLTIMA MODIFICAÇÃO DE DADOS)** - SUFIXO (caso exista mais de uma planilha no referido Programa/Projeto).

* Sendo assim, o nome do arquivo deve adotar o seguinte padrão:

Ex.: **5RC CODIGO DO PROGRAMA/PROJETO NOME DA PLANILHA** **(200114)-1**;

* Em planilhas com especificidades, como dados exclusivos de algum componente, pode-se definir o nome da mesma de acordo com o que é apresentado

Ex.: "MONITORAMENTO DE QUELÔNIOS", ou "SEMENTES", ou "EDIFICAÇÕES", ou "LIVRO DE REGISTRO", entre outros;

* Planilhas diferentes do mesmo Programa/Projeto devem ser individualizadas com sufixos numéricos "-1", "-2", "-3", etc., após o número do Programa/Projeto

Ex.: **5RC CODIGO DO PROGRAMA/PROJETO NOME DA PLANILHA** **(200114)-1**; **5RC CODIGO DO PROGRAMA/PROJETO NOME DA PLANILHA** **(200114)-2**;

* Os arquivos originalmente em EXCEL a serem importados devem estar no formato “.xlsx”.

Ex.: **5RC CODIGO DO PROGRAMA/PROJETO NOME DA PLANILHA** **(200114)-1**.xlsx

# 3 – PLANILHA ÚNICA

Caso não seja possível utilizar apenas uma base de dados, **procure manter uma variável chave que permita vincular as informações das diferentes bases de dados**.

Interface gráfica do usuário, Aplicativo, Tabela, Excel

Descrição gerada automaticamente

# 4 – IDENTIFICADOR

A variável “Identificador” deve ser colocada na primeira coluna da planilha e tem como objetivo discriminar as unidades amostrais (individualizar cada linha).

* Se isso já ocorre na planilha, não haverá a necessidade de implementar novamente;
* Se isso não ocorrer, uma primeira coluna deve ser adicionada, denominada como "Identificador", e os campos devem ser preenchidos com números em sequência, que acompanharão a inserção de novos dados na planilha;
* Essa sequência não será modificada com a inserção de novos dados ao longo do tempo, ela será fixa e novos dados incorporados receberão, sequencialmente, nova numeração.
* Se estiver digitando os dados de um formulário impresso, escreva no questionário físico o número do ID correspondente na base de dados para facilitar futuras consultas ou verificações.

Interface gráfica do usuário, Aplicativo, Tabela, Excel

Descrição gerada automaticamente

# 5 – VARIÁVEIS POR COLUNAS

Preencha o nome de cada variável na primeira linha das colunas seguintes. O importante aqui é a criação de uma variável para cada pergunta do questionário. No exemplo da imagem abaixo, todas informações sobre o estado civil dos entrevistados serão preenchidas na coluna “E” da planilha.

A dúvida que surge neste ponto é: “Como colocar em apenas uma coluna as respostas para questões de múltipla escolha?“

Uma opção é a de estratificar a questão de maneira com que cada opção de resposta tenha uma coluna específica. Em uma pergunta sobre a utilização da internet por exemplo, o respondente poderia selecionar mais de uma resposta. Para cada uma delas, nomearemos uma coluna na planilha de maneira que facilite sua identificação como pertencente à uma questão de múltipla escolha:

* Opção “Estudar e pesquisar” -> “Util\_Estudar”
* Opção “Ler notícias” -> “Util\_Noticias”

Interface gráfica do usuário, Aplicativo, Tabela, Excel

Descrição gerada automaticamente

# 6 – VARIÁVEIS POR LINHAS

Para as pesquisas de campo onde se utilizam formulários em papel por exemplo, cada questionário receberá um número único de identificação, que será preenchido na coluna da variável ID, e sugere-se que esse número seja marcado no questionário impresso para futuras verificações.

Em seguida basta preencher as células com seus respectivos dados. Observe que preencher as respostas das opções de uma questão de múltipla escolha, pode-se colocar “Sim” para o caso da opção ter sido selecionada e “Não” caso contrário.

Tabela

Descrição gerada automaticamente

Nesse ponto, os dados já estão suficientemente estruturados. No entanto, alguns erros podem ter sido cometidos e poderá ser necessário corrigi-los no ambiente de análise. O próximo passo ajuda a evitar alguns destes erros.

# 7 – CODIFICAÇÃO

Durante o processo de digitação, podem ocorrer alguns erros que precisarão ser corrigidos antes da análise. Ao preencher o estado civil dos respondentes, por exemplo, pode-se acabar digitando “solteiro” para um indivíduo e “solteira” para outro, criando duas categorias diferentes que se referem ao mesmo tipo de estado civil.

Para evitar esse tipo de erro, sugere-se que as variáveis sejam codificadas. Para a questão sobre estado civil, por exemplo, pode-se utilizar a codificação:

* Opção “Solteiro” equivale a “1”
* Opção “Casado” equivale a “2”
* Opção “Separado” equivale a “3”
* Opção “Viúvo” equivale a “4”

Em geral, para questões com opções de resposta “Sim” ou “Não” se utiliza a codificação “1” para “Sim” e “0” para “Não”.

Podemos alterar os nomes das variáveis, o que poderá facilitar o trabalho de leitura do sistema gerenciador de banco de dados que será utilizado para armazenamento. Como exemplo, podemos alterar renomeando as variáveis para “Q1”, “Q2”, “Q3” etc., e para aquelas variáveis que surgiram de uma questão de múltipla escolha, para “Q6\_01″, ¨Q6\_02”, etc., onde Q6 representa a pergunta e os dois números finais a opção de resposta (Exemplo: “Util\_Estudar” para “Q6\_01”).

Tabela

Descrição gerada automaticamente com confiança média

Neste ponto é extremamente indicado que seja utilizada uma outra aba no arquivo para que se registre toda a codificação realizada, para que não exista o risco de se perder alguma informação por faltar a correspondência correta entre o código da base de dados e a resposta dos formulários impressos.

Para isso, sugerimos que utilize:

* a primeira coluna para preencher o código da variável,
* a segunda para identificar a questão do questionário,
* a terceira para registrar a codificação utilizada.

Interface gráfica do usuário, Aplicativo, Tabela, Excel

Descrição gerada automaticamente

# 8 – COORDENADAS

TODAS as coordenadas espaciais devem estar presentes no Fuso 22 sul, Datum SAD69, sistema de projeção UTM, somente em NÚMEROS, sem decimais, vírgulas ou pontos:

* A coordenada X deve ser indicada com SEIS DÍGITOS; e,
* A coordenada Y deve ser indicada com SETE DÍGITOS.

# 9 – DATA E HORA

Na coluna para inserção da DATA, os campos devem ser preenchidos no formato DD/MM/AAAA, na coluna para inserção de HORA, os campos devem ser preenchidos no formato hh:mm:ss.

# 10 – DADOS ACUMULADOS

O Banco de Dados Brutos deve apresentar DADOS ACUMULADOS desde o início da execução do Programa/Projeto, até a data de corte do Relatório Consolidado (RC) que o acompanha. Além disso, devem ser EXATAMENTE os mesmos dados apresentados no RC – nem mais, nem menos.