



Estruturação de um *Data Warehouse* para repartição territorial de desmatamentos no bioma Amazônia e o Cadastro Ambiental Rural

Renato da Silva de Deus¹, Mariela Almeida¹, Carlos Alberto de Carvalho¹

¹Embrapa Territorial

Av. Soldado Passarinho, 303, CEP 13070-115 - Campinas - SP

carlos-alberto.carvalho@embrapa.br,

{mariela.almeida,renato.deus}@colaborador.embrapa.br

Abstract. *Due to the large amount of spatial data on the Amazon biome and the need to understand the relationships between rural properties and deforestation, this article defines the proposal of a geospatial Data Warehouse structure in order to organizing this spatial information and relating deforestation phenomena in the Amazon Biome with the territorial dimension occupied by rural properties in the Rural Environmental Registry (CAR). This work presents the conceptual methods for sanitizing CAR data used to develop scripts for preparing and loading data into the Data Warehouse, creating an adequate structure for statistical calculations and data presentation.*

Resumo. *Devido a grande quantidade de dados espaciais sobre o bioma Amazônia e a necessidade de entender os relacionamentos entre os imóveis rurais e desmatamentos, este artigo define a proposta de uma estrutura de Data Warehouse geoespacial com vistas a organizar estas informações espaciais e relacionar os fenômenos de desmatamento no Bioma Amazônia com a dimensão territorial ocupada pelos imóveis rurais do Cadastro Ambiental Rural (CAR). Apresenta os métodos conceituais de sanitização dos dados do CAR utilizados para o desenvolvimento dos scripts de preparação e carga dos dados no Data Warehouse, criando uma estrutura adequada para cálculos estatísticos e apresentação de dados.*

1. Introdução

Este artigo está inserido no projeto “Inteligência Territorial Estratégica aplicada à identificação, quantificação e qualificação de processos de desmatamento no bioma Amazônia e ao apoio e orientação de políticas públicas e privadas na região

amazônica”, parceria entre a Embrapa Territorial e o Centro Gestor e Operacional do Sistema de Proteção da Amazônia voltado para realizar a qualificação numérica, cartográfica, iconográfica e temática dos processos de desmatamento no bioma Amazônia com vistas a apoiar e orientar políticas públicas no bioma com base em inteligência territorial. Este projeto maior nasceu do contexto complexo inerente ao bioma Amazônia, onde diversas dimensões humanas, territoriais, econômicas, agrárias e ambientais se relacionam, da necessidade de proteção, preservação, regeneração e harmonia entre o desenvolvimento dos diversos grupos humanos, sustentabilidade ambiental e econômica e dos processos de desmatamento que ocorrem na região (EMBRAPA, 2019).

Para apreender esse contexto complexo, é fundamental qualificar os processos de desmatamento, pois eles prosseguem combinando tempo e espaço de forma variada, em toda região, com diversos atores e interesses cuja identificação é necessária à compreensão do fenômeno, tanto para orientar nas definições de estratégias e mecanismos de acompanhamento, fiscalização e controle como apoiar a formulação de políticas públicas diferenciadas com base no contexto territorial da região. Diante dessa complexidade de situações, faz-se necessária a organização dessas informações de forma a realizar análises de contexto territorial. Este artigo apresenta uma proposta de estruturação e análise dessas informações sobre o bioma Amazônia.

A organização do artigo se apresenta da seguinte forma: uma seção teórica que aborda as definições e conceitos sobre o Cadastro Ambiental Rural (CAR) e *Data Warehouse* utilizados para a definição dos métodos. Em materiais e métodos são detalhadas, as fontes de dados, áreas de estudos e os métodos para determinar os relacionamentos entre a repartição territorial de interesse e os desmatamentos, bem como os métodos para a construção e estruturação do *Data Warehouse* utilizando um Sistema Gerenciador de Banco de Dados Relacional com extensão Geoespacial. Na seção de resultados são apresentadas demonstrações parciais dos métodos empregados que serviram de base para preparação e carga na estrutura do *Data Warehouse* também exposto nesta seção.

2. Referencial Teórico

O CAR foi criado para integrar as informações ambientais das propriedades e posses rurais, de forma a compor uma base de dados para controle, monitoramento, planejamento ambiental e econômico e combate ao desmatamento (BRASIL, 2012). A base de dados que abriga o CAR é o Sistema de Cadastro Ambiental Rural - SICAR (SICAR, 2021). Ela é única e atualizada através da inserção do cadastro de um imóvel rural no sistema. Os imóveis rurais podem ser classificados por tamanho (pequeno, médio e grande) a depender da sua quantidade de módulos fiscais. Esta classificação garante ao imóvel rural condições diferenciadas (benefícios, tributação etc.) segundo a sua função social.

O INCRA define módulo fiscal como uma forma de medida em hectares definida para cada município. Essa definição leva em consideração a renda e o tipo da exploração

predominante; outros tipos de exploração secundárias, que sejam significativas para o município; e o conceito de propriedade familiar. A classificação dos imóveis rurais fica então da seguinte maneira, de acordo com a lei nº 8.629, de 25 de fevereiro de 1993: pequeno imóvel (≤ 4 módulos fiscais), médio imóvel (> 4 e ≤ 15 módulos fiscais) e grande imóvel (> 15 módulos fiscais) (INCRA, 2020b).

Para a organização das informações, utilizou-se de referencial teórico sobre a construção de *Data Warehouse* (DW), conforme apresentado a seguir.

A Oracle resume o objetivo e algumas características de um *Data Warehouse* da seguinte maneira, “[...] é um tipo de sistema de gerenciamento de dados projetado para ativar e fornecer suporte às atividades de *business intelligence* (BI), especialmente a análise avançada. Os *data warehouses* destinam-se exclusivamente a realizar consultas e análises avançadas e geralmente contêm grandes quantidades de dados históricos. Os dados em um *data warehouse* geralmente são derivados de uma ampla variedade de fontes, como arquivos de *log* de aplicativos e aplicativos de transações.”(ORACLE, 2021). O cientista da computação William H. Inmon define *Data Warehouse* como “[...] um depósito de dados orientado por assunto, integrado, não volátil, variável com o tempo, para apoiar as decisões gerenciais”(ASTERA, 2020).

Existem diversas abordagens para a construção de um *Data Warehouse*, este artigo apresenta a definida por Ralph Kimball onde a relação entre os conceitos de fatos e dimensões se dão da seguinte maneira: Kimball apresenta a tabela fato como: “Uma tabela de fatos contém as medidas numéricas produzidas por um evento de medição operacional no mundo real. Na granularidade mais baixa, uma linha da tabela de fatos corresponde a um evento de medição e vice-versa.”(KIMBALL, 2013). Já as dimensões, como cita Astera, “O esquema em estrela é o elemento fundamental do modelo de *data warehouse* dimensional. Nesse esquema em estrela, uma tabela de fatos é delimitada por várias dimensões. A modelagem dimensional Kimball permite que os usuários construam vários esquemas em estrela para atender a várias necessidades de relatórios. A vantagem do esquema em estrela é que as consultas de pequenas tabelas dimensionais são executadas instantaneamente”(ASTERA, 2020).

Outro processo utilizado para construir um *Data Warehouse* é o ETL (do inglês: Extrair, Transformar e Carregar). A *Spotlessdata*, definiu esse processo como: “O processo de obtenção de dados de seus vários conjuntos de dados e outras fontes para mesclar em um conjunto de dados perfeito em seu *data warehouse* é conhecido como extrair, transformar, carregar (ETL) e é usado ao atualizar partes de sua plataforma ou transferir de um legado para uma nova plataforma. Os dados são extraídos de várias fontes, transformados em dados de qualidade consistente antes de serem carregados no *data warehouse*” (SPOTLESSDATA, 2017).

3. Materiais e Métodos

A seguir são apresentados os materiais utilizados, em fontes de dados e área de estudo, e os métodos e procedimentos aplicados sobre esses conjuntos de dados para a definição dos relacionamentos entre o fenômeno de desmatamento e a dimensão territorial dos Imóveis cadastrados no CAR e da estrutura do *Data Warehouse*. Em linhas gerais, foram utilizados os conceitos pesquisados e expostos acima que mostram que é útil nessa

situação de grande conjunto de dados (desmatamentos, imóveis rurais do CAR, áreas públicas e bioma)

3.1. Fontes de Dados Utilizadas

Os dados vetoriais utilizados vieram das seguintes fontes: 1) Sistema Nacional de Cadastro Ambiental Rural (SICAR); 2) Instituto de Colonização e Reforma Agrária (INCRA); 3) Instituto Brasileiro de Geografia e Estatística (IBGE); 4) Ministério do Meio Ambiente (MMA); e 5) Instituto Nacional de Pesquisas Espaciais (INPE).

Os limites geográficos dos imóveis rurais do CAR foram obtidos a partir de (SICAR, 2021). Por também serem consideradas como imóveis rurais, foram obtidas as unidades de conservação de reservas extrativistas (RESEX) e de desenvolvimento sustentável (RDS) a partir de (MMA, 2020). As informações sobre os assentamentos agrários foram adquiridas no Acervo Fundiário do INCRA (2020a), uma vez que também são considerados como imóveis rurais.

Os limites territoriais foram obtidos na base de dados do IBGE (2020). Para a delimitação do bioma Amazônia utilizou-se a versão disponível em IBGE (2004). Os polígonos de desmatamento foram obtidos junto ao Projeto Monitoramento do Desmatamento da Floresta Amazônica Brasileira por Satélite - PRODES (INPE, 2020). Estes polígonos são gerados a partir de imagens dos satélites e possuem área mínima de 6,25 hectares.

3.2. Área de estudo

O recorte territorial é o Bioma Amazônia. De acordo com IBGE (2019), bioma é uma área de grande extensão geográfica, ocupando 49,5% do território brasileiro, determinada pela vegetação dominante, com condições geoclimáticas semelhantes, onde a fauna e flora vivem de maneira associada. O clima é identificado como quente e úmido, a temperatura média é de 25°C e com chuvas bem distribuídas ao longo do ano. A fitofisionomia predominante é a florestal com vegetação caracterizada como Floresta Ombrófila Densa (Figura 1). A população estimada para o ano de 2017 neste bioma foi de 21 milhões de habitantes, o que corresponde a 10,5% da população do Brasil. O Índice Firjam de Desenvolvimento Municipal (IFDM) de 2016 atribui um valor de 0,568, o que significa desenvolvimento regular (IBGE, 2019).



Figura 1. Mapa de localização do Bioma Amazônia. Fonte: IBGE (2004)

3.3. Métodos e procedimentos

3.3.1. Métodos de definição do relacionamento entre o fenômeno de desmatamento e a dimensão territorial dos Imóveis cadastrados no CAR

Este trabalho envolve 3 etapas sucessivas. A primeira etapa foi realizar uma série de operações de geoprocessamento nos dados brutos do SICAR de 2020, relativas à retirada de possíveis duplicatas de imóveis rurais e separação em 3 categorias de tamanho de imóvel (pequenos, médios e grandes imóveis rurais). Também foi necessária uma atenção especial aos polígonos que representam os perímetros de assentamentos da reforma agrária, uma vez que eles, por definição, possuem sobreposições entre o perímetro do projeto de assentamento e cada um dos lotes que o compõem (Figura 2).



Figura 2 . Imóvel rural do perímetro do projeto do assentamento agrário (em destaque de contorno vermelho) e vários imóveis menores (ou lotes) em seu interior.

A título de esclarecimento, a seção I do capítulo V da Instrução Normativa N° 2/MMA, de 06 de Maio de 2014 descreve sobre os parâmetros dos quais os Assentamentos de Reforma Agrária precisam se enquadrar para efetivação da inscrição no Sistema de Cadastro Ambiental Rural (MMA, 2014). A normativa explica que o cadastramento é de responsabilidade do órgão fundiário e, tanto o perímetro do assentamento quanto os lotes individualizados, devem ser cadastrados no SICAR.

Procedeu-se em retirar as áreas de Projeto de Assentamento (PA) para que fosse evitada a superestimação da contagem de área e quantidade de imóveis rurais do CAR. Essa superestimação ocorre quando um polígono do Projeto de Assentamento – PA tem em seu interior vários polígonos menores, representando os lotes que o compõem, conforme apresentado na Figura 2. Pode ocorrer na base de dados polígonos iguais e sobrepostos. Eles também foram levados em consideração e retirados pela mesma razão, uma vez que essas áreas sobrepostas eram duplicatas de polígonos já existentes.

Por essa ocorrência de projetos de assentamento e de duplicatas acontecerem

mais em polígonos muito grandes, o processo foi realizado para imóveis rurais classificados em grandes e médios e somente aqueles polígonos de imóveis que foram declarados como projetos de assentamento (código AST) no *shapefile* do CAR. O procedimento deu-se pelo cruzamento entre as camadas de assentamento (código da modalidade: PA) do INCRA, Unidades de conservação de categoria extrativista e desenvolvimento sustentável (MMA, 2020) e polígonos grandes do próprio *shapefile* do CAR.

A segunda etapa refere-se à classificação geral entre imóveis predominantemente públicos e predominantemente privados. Esta classificação preliminar e aproximada foi feita para se ter uma ideia de como está a distribuição dos imóveis rurais dentro do CAR em função de entes majoritariamente públicos (como as unidades de conservação do tipo RESEX e RDS) e de entes majoritariamente privados (fora das áreas de RESEX e RDS). Essa aproximação levou em consideração o que se é declarado no CAR e sua intersecção dos imóveis rurais com áreas públicas. Este procedimento foi realizado para pequeno, médio e grande imóvel rural.

A terceira etapa deste processo foi a realização da seleção dos polígonos de desmatamento e dos imóveis rurais que contém os polígonos de desmatamento. Destacam-se duas situações: 1) neste primeiro momento, para os dados de desmatamento não será aplicada a ferramenta de geoprocessamento *clip (recortar)* para que seja evitada a duplicata de informações; 2) assumiu-se os imóveis rurais presentes no CAR de 2020 como já existindo no período entre 2008 e 2019. A ferramenta aplicada foi o *select by location* do software *ArcGIS* (ESRI, 2021).

As 3 etapas descritas acima como operações feitas isoladamente com o Sistema de Informação Geográfica (SIG) local, foram utilizadas como base conceitual para a definição da estrutura, preparação, sanitização e carregamento dos dados no *Data Warehouse*.

Para expressar computacionalmente essas situações de categorias de imóveis e capacidade de realizar análises espaciais, seguiu-se o estudo do desenho mais adequado para a estruturação do *Data Warehouse* que melhor representasse o relacionamento entre o fenômeno de desmatamento e os imóveis rurais do CAR.

3.3.2. Métodos para estruturação do *Data Warehouse* para Repartição Territorial de Desmatamentos no Bioma Amazônia e o Cadastro Ambiental Rural

O objetivo do *Data Warehouse* (DW) é centralizar os dados do projeto provenientes de diversas fontes (*shapefiles* de polígonos de desmatamento, limites dos imóveis rurais do CAR, limites de áreas públicas em unidades de conservação entre outros) e classificá-los de acordo com suas dimensões temporais e territoriais e estabelecer inter-relações sobre essas informações.

O DW facilita as consultas para extração e análise de dados pela forma como organiza a informação em suas estruturas internas. Possibilita realizar cruzamento das informações para responder os questionamentos existentes de maneira rápida e eficiente, pois o banco de dados possibilita não só a conexão com a plataforma *web*, mas com diferentes ferramentas de análises de dados, facilitando a geração de relatórios, gráficos, tabelas e mapas.

4. Resultado

Como resultados, apresentam-se uma demonstração parcial das 3 etapas apresentadas no item anterior e na sequência a estrutura do Data Warehouse. Uma vez com os polígonos de desmatamento que intersectam os imóveis rurais do CAR, foi possível verificar quanto efetivamente esses polígonos de desmatamento estão inseridos nos imóveis rurais. Para isso, a ferramenta *clip* do *ArcGIS* (ESRI, 2021) foi utilizada, de forma a não comprometer a quantidade de polígonos, mas captar a quantidade de área de desmatamento interna ao imóvel rural. Na Figura 3 pode-se ver um polígono de desmatamento dividido entre 3 imóveis rurais (a) e uma parte do polígono de desmatamento não coincidente com imóveis do CAR (b). A quantidade de polígonos foi preservada (1 desmatamento pela ferramenta *select by location* do *software ArcGIS*) e a área efetiva que ocorreu dentro de imóveis rurais (c) ficou mais próxima da realidade, ao descontar a parte em verde (fora de imóveis rurais).

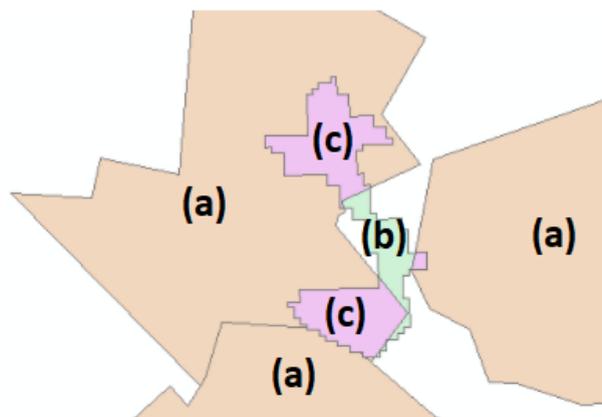


Figura 3. Interseção de um polígono de desmatamento (c)(b) com imóveis rurais (a) após aplicação da ferramenta *clip*.

Para atender as definições dadas no relacionamento entre o fenômeno de desmatamentos e os territórios dos imóveis rurais do CAR desenhou-se o DW utilizando-se do Método Kimball (ASTERA, 2020) com um DW desnormalizado, em esquema estrela (com tabelas fatos e dimensões), definindo-se que o "fato" ou "assunto" (termos da estrutura do DW) seria o polígono de desmatamento, utilizando os dados de "tamanho da área desmatada em hectares" e o "polígono" como atributos. A partir desse "fato", definiram-se as dimensões: territoriais e temporais. Essa estrutura permite o incremento de dimensões conforme as necessidades do projeto. Outra vantagem do conceito de "fato", utilizado pelo *Data Warehouse*, é a possibilidade de identificar quais os territórios fazem intersecção em um único polígono de desmatamento.

Foram utilizados os conceitos de ETL – *Extract, Transform, Load* (Extrair, Transformar e Carregar) para criação dos scripts responsáveis pela extração e carregamento em uma base de dados relacional dos dados referente ao bioma Amazônia e aos imóveis rurais do CAR, além das áreas das Unidades Federativas e dos Municípios, todos presentes em arquivos do tipo *shapefile* e dos *scripts* responsáveis pela transformação e carregamento no *Data Warehouse*.

Tais scripts foram desenvolvidos com base nas definições de classificação e

organização dos dados espaciais do CAR com dois objetivos: 1) processar e armazenar os dados em um banco de dados relacional geoespacial para recuperação dos dados, execução de análises e apresentação em uma plataforma *GeoWeb*; 2) automatização da tarefa para posteriores atualizações dos dados de desmatamento e dos dados do CAR nos anos seguintes.

Por exemplo, uma área de desmatamento em uma região fronteiriça de divisões político-administrativas pode afetar um ou mais Estados da Federação, um ou mais municípios e um imóvel rural cadastrado no Cadastro Ambiental Rural (CAR) ao mesmo tempo (Figura 5). A estrutura desenhada no DW permitiu identificar quais são as áreas que fazem essas intersecções individualmente para cada polígono de desmatamento (ou “fato” do DW). Dessa forma, aumenta a capacidade de qualificação desses desmatamentos, tanto de forma individual quanto no conjunto.

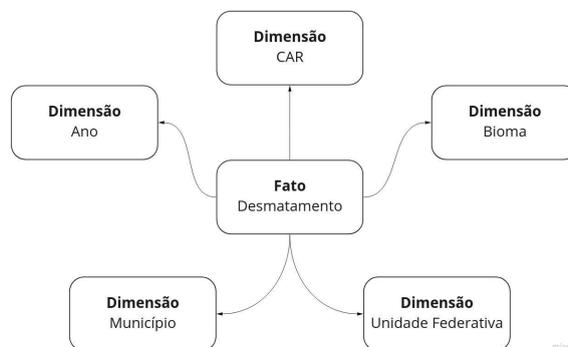


Figura 5. Desenho do *Data Warehouse* em esquema estrela, com a tabela fato ao centro e as dimensões ao redor.

5. Conclusão

A partir dessa proposta, o *Data Warehouse* permitirá a possibilidade de qualificar os polígonos de desmatamento de acordo com suas intersecções, dando visibilidade dos diferentes territórios que um mesmo polígono pode estar vinculado, além dos imóveis rurais do CAR. Também possibilitará agilidade nas consultas dos dados, na geração de estatísticas, relatórios e apresentações em uma plataforma *GeoWeb*.

Ao se trabalhar com os dados de imóveis rurais do CAR é necessário observar e detalhar como estes dados espaciais representam a informação de seus limites geográficos. A classificação de tamanho de módulo fiscal atribuída aos imóveis rurais, assim como a identificação dos polígonos duplicados, necessitou tanto de operações de geoprocessamento quanto uma navegação visual sobre os polígonos. Em algumas vezes, a execução de algumas funções de geoprocessamento presentes no *ArcGIS* (*clip*, *select by location*) precisou de ser acompanhada de uma identificação visual e pontual, pois alguns cruzamentos e intersecções podem não ocorrer de forma como esperada, necessitando de um olhar mais apurado em situações de um polígono apenas tocar o outro.

Conclui-se portanto que os métodos utilizados se mostraram complementares, pois foi possível transpor o trabalho executado em uma estrutura, com ferramentas específicas de geoprocessamento e arquivos *shapefile*, para uma estrutura de *Data*

Warehouse mantendo a relação do fenômeno de desmatamento e das dimensões territoriais dos imóveis rurais do CAR e com a possibilidade de extensão de outras dimensões na mesma estrutura. E com as definições de scripts que automatizam as tarefas e carregam o DW de forma incremental para posterior consumo desses dados nas análises. Todo o material produzido, *scripts* de construção do banco, de sanitização, preparação e carregamento dos dados estão passando por revisões e refinamento para serem disponibilizados ao público ao término do projeto ao qual estão inseridos.

6. Referencias

ORACLE. **O que é um Data Warehouse: Data Warehouse Definido.** 2020.

Disponível em: <https://www.oracle.com/br/database/what-is-a-data-warehouse/>.

Acesso em: 22 out. 2021.

KIMBALL. **Kimball Dimensional Modeling Techniques: Basic Fact Table Techniques.** 2013. Disponível em

<https://www.kimballgroup.com/wp-content/uploads/2013/08/2013.09-Kimball-Dimensional-Modeling-Techniques11.pdf>. Acesso em 22 out. 2021.

ASTERA. **Data Ware House Concepts: Kimball vs. Inmon Approach.** 2020.

Disponível em: <https://www.astera.com/type/blog/data-warehouse-concepts/>.

Acesso em: 20 jun. 2021.

SPOTLESSDATA. **Exploring data warehouses and data quality: Extract, transform, load.** 2017. Disponível em:

<https://web.archive.org/web/20180726071809/https://spotlessdata.com/blog/exploring-data-warehouses-and-data-quality>. Acesso em 22 out. 2021.

BRASIL. **Lei nº 12.651, de 25 de maio de 2012.** Disponível em:

http://www.planalto.gov.br/ccivil_03/ Ato2011-2014/2012/Lei/L12651.htm.

Acesso em: 22 jun. 2021.

EMBRAPA. Sistema de Inteligência Territorial Estratégica do Bioma Amazônia.

2019. Disponível em: <https://www.cnpm.embrapa.br/projetos/siteamazonia/>

ESRI. **ArcGIS 10.7.** 2021. Disponível em: <http://www.esri.com/software/arcgis/arcgis-for-desktop>. Acesso em: 20 jun. 2021.

IBGE. **Biomias.** 2004. Disponível em:

<https://www.ibge.gov.br/geociencias/cartas-e-mapas/informacoes-ambientais/1584-2-biomias.html?edicao=16060&t=sobre>. Acesso em: 12 jun. 2021.

IBGE. **Biomias e sistema costeiro-marinho do Brasil:** compatível com a escala

1:250000. 2019. <https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=2101676>. Acesso em: 20 jun. 2021.

IBGE. **Estrutura Territorial.** 2020. Disponível em <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/estrutura-territorial.html>. Acesso em 22/06/2021.

INCRA. **Acervo fundiário.** 2020a. Disponível em

<http://acervofundiario.incra.gov.br/acervo/acv.php>. Acesso em: 12 jun. 2021.

INCRA. **Módulo Fiscal.** 2020b. Disponível em

<https://antigo.incra.gov.br/pt/modulo-fiscal.html> .html. Acesso em: 20 jun. 2021.

INPE. **TerraBrasilis - Incremento anual no desmatamento - Shapefile**

(2008/2018). 2020. Disponível em

<http://terraBrasilis.dpi.inpe.br/download/dataset/>

[legal-amz-prodes/vector/yearly_deforestation_2008_2018.zip](http://terraBrasilis.dpi.inpe.br/download/dataset/legal-amz-prodes/vector/yearly_deforestation_2008_2018.zip). Acesso em: 20 jun. 2021.

MMA. **Unidades de Conservação**. 2020. Disponível em:

<https://antigo.mma.gov.br/areas-protegidas/cadastro-nacional-de-ucs/dados-georeferenciados.html>. Acesso em: 20 jun. 2021.

MMA. **Instrução Normativa N° 2/MMA, de 06 de Maio de 2014**. Disponível em:

https://www.car.gov.br/leis/IN_CAR.pdf. Acesso em: 20 jun. 2021.

SICAR. **Sistema de Cadastro Ambiental Rural**. 2021. Disponível em

<https://www.car.gov.br/publico/imoveis/index>. Acesso em: 20 jun. 2021.