# **Text S3: Description of deforestation and soy datasets and processing.**

The first dataset is PRODES/INPE (2014). PRODES is an Amazon deforestation monitoring system developed by the National Institute for Space Research (INPE). It produces annual datasets with deforestation increments between August 1st of one year and July 31st of the subsequent year. For example, the 2014 dataset shows all identified polygons with deforestation between August 2013 and July 2014[[1]](#footnote-1). Analogical maps of deforestation were produced between 1988 and 2000, with digital maps starting from 2001 onwards. Nonetheless, digital maps with accumulated deforestation up to 1997, another with accumulated deforestation between 1998 and 2000, and annual maps from 2001 onwards have already been developed (PRODES/INPE, 2014). The second dataset used is PMDBBS/IBAMA (2009), particularly the information on accumulated deforestation in the Cerrado up to 2002. For deforestation in the Cerrado from 2003 onwards, SIAD[[2]](#footnote-2) data from LAPIG/UFG (2014) was used. With these three datasets, it was possible to have a combined dataset with accumulated deforestation in the Amazon up to 1997, accumulated clearing between 1998 and 2000 and yearly increments from 2001 to 2014 (PRODES/INPE, 2014), together with accumulated conversion in the Cerrado up to 2002 (PMDBBS/IBAMA, 2009) and yearly deforestation between 2003 and 2014 (LAPIG/UFG, 2014).

For PRODES/INPE (2014) data, all datasets showing annual deforestation increments were acquired from INPE’s website, including accumulated deforestation up to 1997 and accumulated deforestation between 1998 and 2000 (digital). All those datasets were transformed from vector to raster with cell size of one hectare and joined together to produce a single raster dataset with all deforestation categories. In addition, vector data sources with annual deforestation in the Cerrado between 2003 and 2014 (LAPIG/UFG, 2014) were merged and rasterized. The vector dataset with accumulated deforestation in the Cerrado up to 2002 (PMDBBS/IBAMA, 2009) was also rasterized. Then all those raster files were summed, composing a resulting raster dataset with both annual and accumulated deforestation for the Cerrado (accumulated up to 2002 and annual increments for 2003-2014) and for the Amazon (accumulated up to 2000 and annual increments for 2001-2014). This raster dataset was finally limited to cover only Mato Grosso.

For soy data, a raster map by Macedo et al. (2012) was used. They actually produced annual maps of planted soy between 2001 and 2010, but only the 2010 map was used in order to identify properties that had soy planted in 2010 as the objective was not to analyze soy expansion. This raster dataset was transformed into vector and a spatial merge was carried between the CAR-LAU-INCRA-TL property boundary dataset and the 2010 soy dataset. This identified all properties with soy in 2010. This way, 4,598,030 hectares of soy were identified. According to data from the Association of Brazilian Soy Producers (Aprosoja, 2014), in 2010 there were 6,411,940 hectares of soy planted in Mato Grosso. In this way, 71.7% of the soy planted in Mato Grosso was identified (Tables S3, S4).

1. Some polygons are not identified in one year due to clouds, but it is likely to be captured in subsequent years. [↑](#footnote-ref-1)
2. Integrate System for Deforestation Alerts. [↑](#footnote-ref-2)