

# Análise FAIR dos repositórios de dados meteorológicos, poluentes atmosféricos e dados de fluxo na Floresta Amazônica

Felipe V. de Almeida<sup>1</sup>, Renato O. Miyaji<sup>1</sup>, Lucas O. Bauer<sup>2</sup>,  
Pedro L. P. Corrêa<sup>1</sup>, Luciana V. Rizzo<sup>2</sup>

<sup>1</sup>Escola Politécnica – Universidade de São Paulo (USP)

<sup>2</sup>Universidade Federal de São Paulo (UNIFESP)

{felipe.valencia.almeida, re.miyaji, pedro.correa}@usp.br

{bauer, lrizzo}@unifesp.br

**Abstract.** *The situation of the Amazon Rainforest is in constant debate in the scientific community due to its relevance both in the climatic context and in the national biodiversity. Analysis of the data collected in the region is seen as a powerful resource not only to identify and mitigate anthropic action but also to understand the complex ecosystem that exists. This article presents an analysis following the FAIR principles of two databases available on the internet and which have data related to climatic measurements of the Amazon Rainforest.*

**Resumo.** *A situação da Floresta Amazônica está em constante debate na comunidade científica devido a sua relevância tanto no contexto climático quanto na biodiversidade nacional. A análise dos dados coletados na região é vista como poderoso recurso não somente para identificar e mitigar a ação antrópica, mas também para entender o complexo ecossistema existente. Este artigo apresenta uma análise seguindo os princípios FAIR de dois bancos de dados disponíveis na internet e que possuem dados relacionados a medições climáticas da Floresta Amazônica.*

## 1. Introdução

É perceptível a importância da Floresta Amazônica tanto em um cenário nacional quanto global. Frequentemente são veiculadas notícias relacionadas ao seu estado, com enfoque na questão climática e no seu desmatamento.

A agenda da ONU 2030 elencou 17 Objetivos de Desenvolvimento Sustentável (ODS), sendo que deste, o objetivo número 13 - Ação contra a mudança global do clima possui relação direta com sua preservação [IBGE 2022]. Observando de maneira aprofundada este objetivo, percebe-se que ele é composto por 8 indicadores, onde é importante destacar o indicador 13.3 - Melhorar a educação, aumentar a conscientização e a capacidade humana e institucional sobre mitigação, adaptação, redução de impacto e alerta precoce da mudança do clima. Este indicador, em pesquisa realizada no dia 15/02/22, consta como “Sem dados”, apontando um horizonte nebuloso para o seu cumprimento. Entende-se que o estudo e análise de dados da Floresta Amazônica pode auxiliar na mitigação dos efeitos da ação antrópica, auxiliando na tomada de decisões e em sua preservação.

Este trabalho apresenta uma análise segundo os princípios FAIR de repositórios de dados climáticos da região da Floresta Amazônica. Inicialmente é feita uma introdução

teórica definindo o que são estes princípios e quais são seus componentes. Em sequência são apresentados dois repositórios contendo diferentes dados climáticos da Floresta Amazônica seguido das considerações finais. Espera-se que a análise aqui realizada possa auxiliar futuros pesquisadores que tenham como propósito realizar projetos que necessitem destes dados.

## 2. Princípios FAIR

Em 2016 foi publicado o guia de princípios FAIR, na revista *Scientific Data* [Wilkinson et al. 2016]. O acrônimo FAIR vem de *Findability, Accessibility, Interoperability and Reusability*, ou seja, é um conjunto de atributos desejáveis que um *dataset* disponibilizado possua, visando atender interesses de diversos *stakeholders*, sendo eles pertencentes à indústria ou a academia. Esses interesses estão relacionados a uma necessidade crescente em se adotar uma postura crítica sobre a disponibilidade dos dados. Trabalhos na área de Ciência de Dados rotineiramente dependem de grandes repositórios de dados, onde pesquisadores podem obter *datasets* específicos para o escopo do seu trabalho. A dificuldade na obtenção destes dados está relacionada a más práticas, podendo neste caso ser mitigadas com a aplicação dos princípios FAIR. Destaca-se também a vantagem da adoção destes princípios na questão da reprodutibilidade de experimentos, que é um problema atual identificado na literatura [McNutt 2014] [Peng 2015].

A seguir é feita uma breve descrição de cada um dos princípios FAIR, apresentando também os tópicos que compõem cada princípio em tradução livre. A versão original destes tópicos pode ser obtida em <https://www.go-fair.org/fair-principles/>. Ressalta-se que a ordem em que os princípios são apresentados está relacionada com a ordem natural de um experimento da área de Ciência de Dados, ou seja, encontrar o dado, acessar o dado, integrar o dado com outras fontes/ferramentas e, por fim, o reuso deste.

*Findability* pode ser traduzido para “encontrabilidade” sendo neste contexto a facilidade com que determinado *dataset* pode ser obtido. *Datasets* disponibilizados em repositórios de grande relevância para a comunidade científica que vão de encontro a este princípio. Além disso, os metadados podem auxiliar na indexação do *dataset* em fontes de pesquisa, facilitando com que este seja encontrado.

Os tópicos relacionados a este princípio são:

- F1. (Meta)dados possuem um identificador global único e persistente
- F2. Dados são descritos com metadados de maneira abundante
- F3. Metadados incluem de maneira clara e explícita o identificador dos dados que eles descrevem
- F4. (Meta)dados são registrados ou indexados em uma fonte de pesquisa

*Accessibility* é a “acessibilidade” ou seja, o quão acessível o *dataset* está uma vez que este é encontrado no passo anterior. Neste princípio entram conceitos de protocolo de comunicação e autenticação, porém é dado aqui um enfoque na necessidade do conjunto de metadados possuir alta acessibilidade, mesmo quando o *dataset* atrelado não estiver acessível.

Os tópicos relacionados a este princípio são:

- A1. (Meta)dados são recuperáveis pelo seu identificador por um protocolo de comunicação padronizado
- A1.1. O protocolo é aberto, gratuito e universalmente implementável
- A1.2. O protocolo permite autenticação e um procedimento de autorização quando necessário
- A2. Os metadados são acessíveis mesmo quando os dados não estão disponíveis

*Interoperable* é a “interoperabilidade”, termo este comumente utilizado para tratar a comunicação entre sistemas distintos. No contexto de dados, este princípio indica a capacidade do dado ser integrado com outras fontes de dados, além de seu uso em ferramentas diversas.

Os tópicos relacionados a este princípio são:

- I1. (Meta)dados usam uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento
- I2. (Meta)dados usam vocabulários que seguem os princípios FAIR
- I3. (Meta)dados incluem referências qualificadas para outros (meta)dados

*Reusability* ou “reusabilidade” é o propósito final dos princípios FAIR, ou seja, o *dataset* gerado a partir de um experimento da área de Ciência de Dados poder eventualmente ser reutilizado em outros experimentos, gerando assim um ciclo completo. Neste princípio os metadados assim como a linguagem utilizada possuem papel fundamental em sua garantia.

Os tópicos relacionados a este princípio são:

- R1. (Meta)dados são descritos de maneira abundante com uma pluralidade de atributos precisos e relevantes
- R1.1. (Meta)dados são disponibilizados com uma licença clara e acessível de uso dos dados
- R1.2. (Meta)dados estão associados com a uma proveniência detalhada dos dados
- R1.3. (Meta)dados atendem aos padrões de comunidade relevantes ao domínio

### **3. Análise dos Dados**

Nesta seção são apresentados dois repositórios de dados climáticos relacionados à Floresta Amazônica. Em cada repositório é delimitado seu escopo tanto em relação ao volume de dados, quanto na questão temporal (período em que os dados foram coletados) e espacial (área coberta pelas medições), além da análise segundo os princípios FAIR.

#### **3.1. Experimento GOAmazon 2014/15**

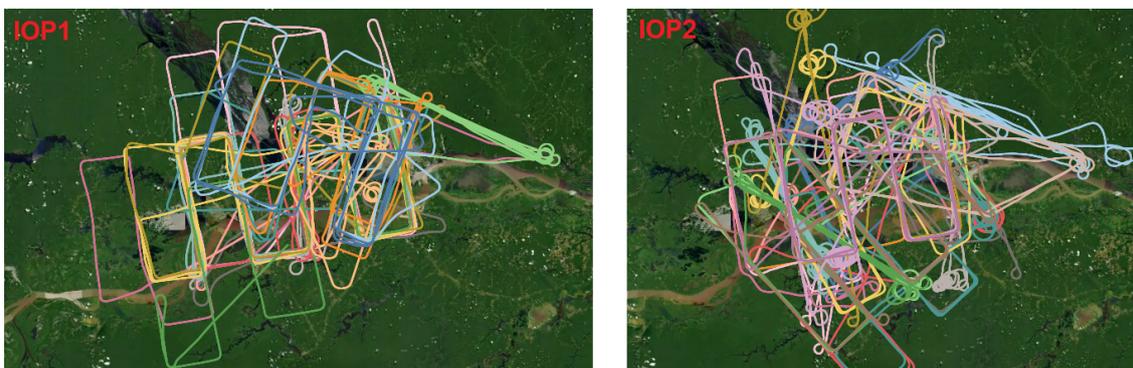
Entre os anos de 2014 e 2015, o *Atmospheric Radiation Measurement* (ARM) - órgão que possui vínculo com o Departamento de Energia dos Estados Unidos da América (*Department of Energy* - DoE) - conduziu o experimento *Green Ocean Amazon* (GOAmazon 2014/15). Esse teve dois anos de duração e realizou coletas de dados meteorológicos e relacionados a aerossóis na região da Bacia Amazônica, mais especificamente nas proximidades da metrópole de Manaus (AM). Essa está situada na confluência do rio Solimões com o rio Negro, que formam o rio Amazonas.

A escolha desta localização específica se deu por se caracterizar como um laboratório com as características ideais para aprimorar a compreensão a respeito do funcionamento dos ciclos químicos, hidrológicos, energéticos e de ecossistemas que se desenvolvem na maior floresta tropical do mundo. Ademais, com os dados coletados, também buscou-se entender como esses serão afetados em cenários com níveis de estresse ambiental e poluição maiores [Martin et al. 2016].

Devido às diferentes condições que se desenvolvem na região Amazônica durante as estações seca e chuvosa, o experimento foi dividido em dois períodos de operação intensiva. O primeiro ocorreu na estação chuvosa durante os dias primeiro de fevereiro e 31 de março de 2014, ao passo que o segundo foi realizado entre os dias 15 de agosto e 15 de outubro do mesmo ano [Martin et al. 2016].

Para a coleta dos dados foram utilizadas tanto estações de pesquisa fixas, quanto móveis. As primeiras representavam nove pontos de coleta dispostos ao longo da região de interesse, englobando também instalações científicas previamente existentes. Duas delas estavam localizadas na cidade de Manacapuru (AM), três em Manaus (AM) e o restante em regiões florestais, como o *Amazon Tall Tower Observatory* (ATTO).

Já as estações móveis foram duas aeronaves pertencentes ao DoE, que realizaram voos de baixa altitude na região de estudo. Foram 35 voos, sendo 16 deles durante a estação chuvosa e 19 na seca. As trajetórias percorridas durante esses voos são apresentadas na Figura 1. Esses foram capazes de abranger uma área maior que a delimitada pelas estações de coleta terrestres, apesar de estarem restritos a curtos períodos de tempo. Assim, enquanto os dados advindos das estações de pesquisa fixas possuem uma resolução temporal maior, na ordem de minutos, os coletados pelas aeronaves apresentam uma resolução espacial superior.



**Figura 1. Trajetórias percorridas pelas aeronaves do experimento GOAmazon 2014/15 durante a) estação chuvosa e b) estação seca**

Os dados coletados pelo experimento e provenientes das estações de pesquisa fixas e móveis são disponibilizados no repositório de dados do ARM [ARM 2022]. Nele, são apresentados mais de 65 *datasets* distintos referentes ao experimento GOAmazon 2014/15. Desses, foram selecionados quatro para serem detalhados e analisados a partir dos princípios FAIR.

O *dataset* [Artaxo 2015] apresenta os dados coletados por uma das estações de pesquisa terrestre, localizada nas coordenadas -3,1392 de latitude e -60,1315 de longitude, nas proximidades da cidade de Manaus (AM), entre os dias primeiro de janeiro e

31 de dezembro de 2014. A resolução temporal adotada é de 5 e 30 minutos. São disponibilizadas medições de nove variáveis de aerossóis distintas, como a concentração de monóxido de carbono, de dióxido de nitrogênio, entre outros.

Já nos *datasets* [Mei 2014], [Mei et al. 2014] e [Butterfield 2015] são disponibilizados os dados referentes aos voos de uma das aeronaves do DoE durante os dois períodos de operação, ou seja, entre primeiro de fevereiro e 31 de março e entre 15 de agosto e 15 de outubro de 2014. Esses foram realizados entre o final da manhã e o início da tarde no horário local. A área compreendida pelos voos é a mesma apresentada na Figura 1, ou seja, percorrendo a região entre as cidades de Manaus (AM) e Manacapuru (AM). Em [Mei 2014] são disponibilizados os dados de concentração numérica de partículas e pressão atmosférica. Já em [Mei et al. 2014] são apresentadas mais de 46 variáveis, sendo a maior parte delas relacionadas às condições de voo da aeronave. Em [Butterfield 2015], as medições de concentração de poluentes e aerossóis estão disponíveis.

A partir dos princípios FAIR, esses *datasets* podem ser avaliados. Em relação à *Findability*, todos os *datasets* cumprem esse requisito por estarem disponibilizados nos repositórios do ARM. Entretanto, apenas [Mei 2014] e [Mei et al. 2014] apresentam metadados abundantes, com descrições detalhadas a respeito das coletas dos dados e de cada variável apresentada. A respeito de *Accessibility*, os *datasets* também satisfazem essa condição, uma vez que estão acessíveis por meio de um protocolo livre, gratuito e que permite autenticações e exige permissões. Além disso, os metadados mostram-se disponíveis mesmo quando os dados não estão.

Já em relação à *Interoperable*, apenas [Mei 2014] e [Mei et al. 2014] possuem metadados com linguagem adequada e que permita a representação completa do conhecimento. Por outro lado, somente no caso de [Artaxo 2015] que é fornecida uma referência qualificada para os metadados. Por fim, sobre *Reusability*, [Mei 2014] e [Mei et al. 2014] satisfazem completamente esse requisito, uma vez que, a partir dos metadados, é possível obter todas as informações necessárias para reutilizar os dados para outros experimentos de Ciência de Dados. Entretanto, ressalta-se que, apesar dos metadados de [Artaxo 2015] e [Butterfield 2015] não estarem completos a partir da abordagem FAIR, ainda assim pode-se reutilizá-los em outros contextos.

### **3.2. Dados das Torres do programa LBA-ECO CD - 32 (ORNL-DAAC)**

O repositório *Distributed Active Archive Center* (DAAC) do laboratório nacional científico dos EUA *Oak Ridge National Laboratory* (ORNL) disponibiliza os dados do programa: Experimento de Larga-Escala da Biosfera-Atmosfera na Amazônia (*Large-Scale Biosphere-Atmosphere Experiment in Amazonia* - LBA). Programa de cooperação científica internacional, com apoio financeiro pátrio e estrangeiro, sob o gerenciamento do Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTI), e, execução do Instituto Nacional de Pesquisas da Amazônia (INPA) [Luizão et al. 2005]. E tem como objetivos entender o funcionamento e o papel da floresta Amazônica em escala regional, e, compreender como as mudanças no clima e no uso e cobertura do solo vão afetar as funções físicas, químicas e biológicas da floresta e o desenvolvimento sustentável da região, mas também a influência da Floresta Amazônica no clima global [Avisar et al. 2002].

O DAAC dispõe uma série de compilações de dados referentes ao programa LBA,

incluindo-se o *Large-Scale Biosphere-Atmosphere Experiment - Ecology Datasets CD 32 Flux Tower Network Data Compilation, Brazilian Amazon: 1999 - 2006*, com dados integrados de projetos independentes de fluxos de superfície desenvolvidos entre 1999 a 2006 com verificação de controle de qualidade e agregação desses dados em diferentes intervalos de tempo (dados horários, dados diários e dados mensais diurnos e noturnos) que contemplam 105 variáveis, tais como fluxos de vórtices turbulentos de carbono e energia, dados de umidade do solo, perfis de dióxido de carbono, radiação e dados meteorológicos de oito torres localizadas na bacia Amazônica e de uma torre localizada no estado de São Paulo.

Ademais, o repositório disponibiliza os dados do *Large-Scale Biosphere-Atmosphere Experiment - Ecology Datasets CD 32 LBA Model Intercomparison Project (LBA-MIP)*, o qual contém observações meteorológicas horárias de temperatura, umidade específica, módulo da velocidade do vento, precipitação, dióxido de carbono, pressão a nível de superfície, radiação de onda curta e radiação de onda longa incidentes à superfície para as 9 torres no período de 1999 a 2006 [de Gonçalves et al. 2013]. Os dados foram padronizados e aplicou-se uma técnica simples de imputação a essa base de dados que serviu como informação de entrada para o Modelo de Intercomparação do LBA (LBA-MIP) [de Gonçalves et al. 2013]. Um resumo de cada torre é apresentado na Tabela 2, informações adicionais podem ser obtidas no site do Centro de distribuição de arquivos de ORNL <sup>1</sup>.

Nesse mesmo DAAC é possível encontrar a versão 2 dos dados de fluxos de carbono e energia das torres, que se trata de uma versão com mais de 200 variáveis, com a inclusão de novas opções, incluindo-se variáveis referentes à capacidade fotossintética, à produtividade bruta da área, respiração, e também, novas opções de variáveis de fluxos de carbono sem dados faltantes e corrigidos para condições atmosféricas de baixa turbulência.

<i>Torres</i>	<i>Coordenadas Geográficas WGS 84</i>	<i>Série Temporal - Flux Tower Network</i>	<i>Série Temporal MIP</i>
<i>TOC_BAN JAV</i>	<i>9.8244 S, 50.1591 O</i>	<i>24/10/2003 a 08/12/2006</i>	<i>01/01/2004 a 31/12/2006</i>
<i>MAN_K34</i>	<i>2.6090 S, 60.2091 O</i>	<i>14/06/1999 a 30/09/2006</i>	<i>01/01/2002 a 31/12/2005</i>
<i>STM_K67</i>	<i>2.8570 S, 54.9590 O</i>	<i>02/01/2002 a 23/01/2006</i>	<i>01/01/2002 a 31/12/2004</i>
<i>STM_K77</i>	<i>3.0202 S, 54.8885 O</i>	<i>01/01/2000 a 30/12/2005</i>	<i>01/01/2001 a 31/12/2005</i>
<i>STM_K83</i>	<i>3.0170 S, 54.9707 O</i>	<i>29/06/2000 a 12/03/2004</i>	<i>01/01/2001 a 31/12/2003</i>
<i>RON_RJA</i>	<i>10.0780 S, 61.9331 O</i>	<i>23/03/1999 a 14/11/2002</i>	<i>01/01/2000 a 31/12/2002</i>
<i>RON_FNS</i>	<i>10.7618 S, 62.3572 O</i>	<i>04/02/1999 a 04/11/2002</i>	<i>01/01/1999 a 31/12/2001</i>
<i>PA_CAX</i>	<i>1.7483 S, 51.4536 O</i>	<i>01/01/1999 a 30/07/2003</i>	<i>01/01/1999 a 31/12/2003</i>
<i>SP_PDG</i>	<i>21.6195 S, 47.6499 O</i>	<i>01/01/2004 a 31/12/2006</i>	<i>01/01/2001 a 31/12/2003</i>

**Tabela 1. Informações referentes às áreas de estudo das torres e série temporal de dados**

As áreas de estudo de cada torre apresentam as seguintes características, conforme descrito por SALESKA et al., (2013).:

- TOC\_BAN JAV: Floresta sazonalmente inundada com a ocorrência de um ecossistema ecótono: bioma cerrado e campo (área de pastagem natural), localizada no estado de Tocantins, na Ilha de Bananal.
- MAN\_K34: Floresta tropical localizada no Km 34 da cidade de Manaus no estado de Amazonas.

<sup>1</sup> [https://daac.ornl.gov/cgi-bin/dataset\\_lister.pl?p=9](https://daac.ornl.gov/cgi-bin/dataset_lister.pl?p=9)

- STM\_K67: Floresta tropical localizada no Km 67, próxima ao município de Santarém no estado de Pará.
- STM\_K77: Região de pastagem e agricultura localizada no Km 77, próxima ao município de Santarém no estado de Pará.
- STM\_K83: Floresta tropical com extração seletiva de madeira localizada no Km 83, próxima ao município de Santarém no estado de Pará.
- RON\_RJA: Floresta tropical seca localizada na Reserva Jarú no estado de Rondônia.
- RON\_FNS: Área de pastagem localizada na Fazenda Nossa Senhora no estado de Rondônia.
- PA\_CAX: Floresta tropical localizada no município de Caxiuana no estado de Pará.
- SP\_PDG: Região com ocorrência do bioma cerrado localizada na Reserva Pé-de-Gigante no estado de São Paulo.

Os dados do LBA-ECO CD-32 disponíveis no repositório DAAC cumprem com os requisitos dos itens de análise FAIR pela facilidade de obtenção dos dados, relevância do repositório para a comunidade científica, descrição dos dados a partir de metadados abundantes e detalhados (*Findability*); acessibilidade aos metadados, mesmo quando os dados não estão disponíveis os metadados são disponíveis. E há um protocolo aberto e gratuito para acessar os dados (*Accessibility*); a possibilidade de integrar os dados com outras fontes de dados, inclusão de referências qualificadas para os metadados, uso de linguagem formal, acessível e aplicável para disseminação do conhecimento (*Interoperable*); e por poder ser reutilizado em diversos experimentos de Ciência de dados (*Reusability*).

#### **4. Conclusão e Trabalhos Futuros**

Este trabalho apresentou uma análise FAIR em dois bancos de dados distintos contendo *datasets* relacionados a medições climáticas na região da Floresta Amazônica. A análise aqui realizada teve como propósito auxiliar futuramente eventuais pesquisadores que desejem realizar projetos tanto na área de Ciências Atmosféricas, com enfoque na análise dos dados, quanto na área de Ciência de Dados, com a possibilidade de treinamento de modelos de aprendizagem de máquina, ou então o uso de métodos de interpolação, como apresentado no estudo de caso.

Como trabalhos futuros pode-se explorar os *datasets* relacionados a dados de satélites. Estes em geral possuem escopo mais amplo, permitindo ao usuário delimitar a área de cobertura dos dados (latitude/longitude), além da data de coleta deles e as variáveis que se deseja obter. Além disso, existe a possibilidade de se construir bases de dados integradas juntando dados de superfície de torres com dados de satélites e de campanhas de amostragem de aviões para a realização de experimentos de Ciência de Dados. Essa união de informações de fontes distintas contribui para a descoberta de novos conhecimentos referentes a temas diversos e complexos. Como exemplo pode-se citar a influência das trocas de carbono entre a biosfera e atmosfera na Floresta Amazônica sob a crescente perturbação antrópica, modelagem de ocorrência de espécies, modelagem da qualidade do ar, dentre outras possibilidades de estudos.

#### **Agradecimentos**

Os autores agradecem ao ARM e as instituições nacionais pela disponibilidade dos dados aqui analisados. Também é feito um agradecimento especial ao projeto FAPESP "Emissão

de gases de efeito estufa na Amazônica e Sistemas de Análise de Dados e Serviços” e ao “Centro de Pesquisa e Inovação de Gases de Efeito Estufa - RCG2I”(2020/15230-5) onde este trabalho está inserido.

## Referências

- ARM (2022). Atmospheric Radiation Measurement User Facility. <https://www.arm.gov/>. Acesso: 20/02/2022.
- Artaxo, P. Barbosa, H. (2015). LFA-USP aerosol data products T2 2014 2015. Atmospheric Radiation Measurement.
- Avissar, R., Silva Dias, P. L., Silva Dias, M. A., and Nobre, C. (2002). The large-scale biosphere-atmosphere experiment in amazonia (Iba): Insights and future research needs. *Journal of Geophysical Research: Atmospheres*, 107(D20):LBA–54.
- Butterfield, Z. Dubey, M. (2015). GoAmazon Aerial Campaign LANL Pircarro IOP1 IOP2. Atmospheric Radiation Measurement.
- de Gonçalves, L., Restrepo-Coupe, N., da Rocha, H., Saleska, S., and Stockli, R. (2013). LBA-ECO CD-32 LBA Model Intercomparison Project (LBA-MIP) Forcing Data. *ORNL DAAC*.
- IBGE (2022). Indicadores Brasileiros para os Objetivos de Desenvolvimento Sustentável. <https://odsbrasil.gov.br/>. Acesso: 20/02/2022.
- Luizão, F. J., Nobre, C. A., and Manzi, A. O. (2005). Projeto LBA: Estudando as complexas interações da biosfera com a atmosfera na amazônia. v.20:p.681–708.
- Martin, S. T., Artaxo, P., Machado, L. A. T., Manzi, A. O., Souza, R. A. F. d., Schumacher, C., Wang, J., Andreae, M. O., Barbosa, H., Fan, J., et al. (2016). Introduction: observations and modeling of the green ocean amazon (goamazon2014/5). *Atmospheric Chemistry and Physics*, 16(8):4785–4797.
- McNutt, M. (2014). Reproducibility.
- Mei, F. (2014). Condensation Particle Counter (CPC) 3010 and 3025(TSI) - GoAmazon, 2014. Atmospheric Radiation Measurement.
- Mei, F., Hubbe, J. M., Comstock, J., Tomlinson, J., and Iwagadts, M. (2014). Data Set from GoAmazon. Atmospheric Radiation Measurement.
- Peng, R. (2015). The reproducibility crisis in science: A statistical counterattack. *Significance*, 12(3):30–32.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, v.3(n.1):p.1–9.