# Compreendendo o processo de indexação, qualificação e associação de dados

Indexação e qualificação dos registros	1
Qualificação dos nomes científicos	1
Qualificação das coordenadas geográficas	1
Exemplo de um registro online indexado e qualificado	2
Uso da Interface de busca e visualização do resultado	2
FILTROS	3
Busca livre	3
Bloco identificação	3
Bloco coleta	4
Bloco coleção	4
Bloco microbial	4
Bloco filtros geográficos	4
Bloco Uso e cobertura da terra MapBiomas	5
Bloco Dados agregados e modificadores	5
Formas de visualizar o resultado de uma busca (outputs)	6
REGISTROS	6
INVENTÁRIOS	7
MAPA	8
IMAGENS	8
NÚMEROS	9
DUPLICATAS	10
BAIXAR	10
COMO CITAR	11
A importância dos filtros de busca	14
Data Cleaning	14
Interface de busca como ferramenta para encontrar possíveis erros	15
Estatísticas de uso dos dados	16
Dashboard	18
API do serviço web v1.0	20
Cadastro do usuário	20

### Interface de busca da Rede speciesLink

specieslink.net/search/

Esse documento tem por objetivo apresentar os principais elementos da interface de busca, recuperação e visualização dos resultados na rede *species*Link.

### cría

dnei 👻 Adm 👻 Sair 🔹 Português

#### species desde 2002 Link



FPS00208 - Callianthe rufinerva

A rede speciesLink promove o acesso livre e aberto a dados, informações e ferramentas para qualquer indivíduo ou grupo, uma vez que os provedores de dados expressaram seu compromisso em compartilhar somente dados não confidenciais. O objetivo da rede é fomentar a pesquisa, a educação e a formulação de políticas para promover a conservação e o uso sustentável da biodiversidade. Os provedores de dados não garantem a precisão dos dados servidos através

da rede *species*Link e seus serviços. Portanto o uso dos dados acessados por meio do portal e serviços web é de responsabilidade do usuário. As restrições indicadas por cada provedor de dados devem ser observadas e respeitadas, e a fonte dos dados deve ser citada de acordo com o texto de citação da rede *species*Link.

busca mapa da rede provedores de dados dashboard uso da rede serviço web como participar blog

### Indexação e qualificação dos registros

Cada registro de ocorrência de espécies enviado à rede *species*Link é submetido a um processo de indexação. Essa indexação tem como objetivo integrar os dados em uma base de dados única para busca e recuperação. Durante o processo de indexação, outras informações são associadas aos registros como **qualificadores do nome científico e coordenadas geográficas**.

### Qualificação dos nomes científicos

O *species*Link utiliza dicionários de nomes científicos para checar e atribuir o status taxonômico aos nomes em cada registro (aceito com autor, aceito sem autor, sinônimo com autor, sinônimo sem autor, ambíguo com autor, ambíguo sem autor, e não encontrado). A checagem de cada nome científico (família, gênero, espécie ou táxon infraespecífico) segue uma lógica para garantir que o status atribuído a ele seja único em toda a rede. Os dicionários são atualizados a cada seis meses e uma nova indexação de todos os registros é realizada. Assim, a indexação condiz com a natureza dinâmica da taxonomia. Os seguintes dicionários são utilizados no *species*Link: *Flora e Funga do Brasil, MycoBank, AlgaeBase, Catálogo de Abelhas Moure, LPSN-List of Prokaryotic names with Standing in Nomenclature e GBIF Backbone Taxonomy*. Uma vez que um status é atribuído a um determinado nome, esse mesmo status é atribuído a todos os registros que tenham o mesmo nome. O status só pode ser alterado quando os dicionários forem alterados e uma nova indexação geral for feita.

### Qualificação das coordenadas geográficas

O sistema analisa e qualifica as coordenadas geográficas de cada registro e indica se são coordenadas **originais**, **consistentes** ou **suspeitas**. Uma coordenada **coordenada suspeita** é aquela que não cai no município indicado. Quando um registro tem o município da coleta e não tem coordenada, o sistema atribui ao registro a coordenada do município associada ao erro máximo. Neste caso o registro recebe um *tag* **coordenada por município**. **Coordenadas** 





**bloqueadas** são aquelas que o curador da coleção bloqueia a transmissão da coordenada geográfica para a rede.

O usuário pode utilizar esses qualificadores na sua busca. A interface de busca possui cinco filtros para a coordenada geográfica: *por município, originais, consistente, suspeitas e bloqueadas*. No caso da busca por coordenadas originais, o usuário poderá selecionar os filtros *consistente* ou *suspeitas*. Assim, o usuário poderá "limpar" o resultado da sua busca ao escolher as coordenadas consistentes, ou o curador poderá recuperar os registros com coordenadas suspeitas para avaliação e correção.

Com a disponibilização aberta dos dados de uso e cobertura da terra do MapBiomas, o CRIA também associa às coordenadas geográficas dos registros de coletas realizadas no Brasil e na PanAmazônia com o histórico da cobertura e uso da terra.

### Exemplo de um registro online indexado e qualificado

A figura a seguir apresenta o registro do espécime UB 227457 com o nome da espécie e as coordenadas geográficas qualificadas.



O registro apresenta *links* para outras fontes de informação utilizadas na qualificação dos dados. O usuário poderá clicar nas setas para obter mais informações como: visualizar a coordenada em um mapa, acessar os dados da espécie na Flora e Funga e no GBIF e passar o mouse no diagrama do MapBiomas que indica que a amostra foi coletada no ano de 2008 em área de formação florestal onde até 2023 não houve mudança no uso e cobertura da terra. A qualificação também indica que se trata de uma espécie ameaçada de extinção criticamente em perigo. O usuário também pode acessar o *link* para a página do herbário UB na rede *species*Link.

### Uso da Interface de busca e visualização do resultado

specieslink.net/search

FILTROS EREGISTROS INVENTÁRIOS 🕈 MAPA 💿 IMAGENS INÚMEROS CO DUPLICATAS 🕹 BAIXAR I COMO CITAR

O menu principal da página de busca da rede *species*Link apresenta ao usuário várias opções de busca e visualização do resultado. A opção **FILTROS** dá acesso aos campos e filtros que o usuário pode utilizar para estruturar a sua busca de acordo com a informação desejada. As



demais opções - **REGISTROS**, **INVENTÁRIOS**, **MAPA**, **IMAGENS**, **NÚMEROS**, **DUPLICATAS** e **BAIXAR** são diferentes formas de visualizar o resultado da busca. A opção COMO CITAR fornece os elementos da busca realizada para uma eventual citação. O sistema sempre destaca o item que está sendo acessado em LARANJA.

### **FILTROS**

Contém os campos disponíveis para restrição do resultado da busca às informções desejadas. São separados nos seguintes blocos: *busca livre, identificação, coleta, coleção, microbial, MapBiomas,* e *filtros geográficos.* Em vários campos é possível especificar mais de um valor, significando que pelo menos um dos valores deve estar presente no resultado. Para isso, pode-se usar vírgula (, ) ou barra vertical ( | ) para separar os valores a serem procurados.

Em alguns campos o usuário encontrará um *i* que poderá fornecer informações específicas sobre o campo. Da mesma forma, no nome do campo tem um x que pode ser utilizado para apagar o conteúdo do campo. Alguns campos apresentam uma flecha direciona o usuário para outra página com campos específicos.

No topo do formulário, à direita, o usuário encontrará a frase **limpar filtros** que, ao ser clicada, limpa todos os campos que foram preenchidos.

#### **Busca livre**

busca livre

Nesse campo, o usuário pode digitar palavras que serão buscadas nos registros de ocorrência dos espécimes. Trata-se de uma busca no texto do registro, portanto não é um campo que oferece uma busca exata. Na busca por *flor amarela*, por exemplo, o sistema irá buscar essas palavras em qualquer campo dos registros enviados à rede. O sistema retornará, por exemplo, registros que têm *"flor amarela"* e que têm frases descritivas como *"flores e botões florais brancos, estames amarelos; frutos imaturos verdes"*. Portanto, a busca livre deve ser utilizada sabendo que o sistema poderá recuperar registros que não atendem o interesse do usuário.

### Bloco identificação

filo ×	classe ×	ordem ×	
nome cie	ntífico i ×		
			2
	filo × nome cie	filo × classe × nome científico i ×	filo × classe × ordem × nome científico i ×

Nesse bloco é possível realizar buscas em campos específicos sobre a taxonomia de um espécime. Trata-se de uma busca com maior precisão. O usuário pode fazer buscas do *Reino* até o *nome científico* da espécie. Além da busca exata é possível selecionar a ferramenta *fonética* para permitir a recuperação de registros com algum erro de grafia. O campo determinador pode também ser problemático uma vez que é possível encontrar diferentes grafias para o mesmo determinador.

#### **Bloco coleta**

filtros geográficos		
países da américa do sul	•	
regiões brasileiras	•	
estados brasileiros	•	
municípios brasileiros	•	
biomas	•	
divisão hidrográfica nacional - macro I	•	
divisão hidrográfica nacional - meso	•	
livisão hidrográfica nacional - micro	•	
unidades de conservação federais	•	
inidades de conservação estaduais	•	
unidades de conservação municipais	•	
erras indígenas	•	
egetação brasileira 2004		
coleta		
coletor ×	número ×	ano da coleta $i$ $ imes$
país ×	estado i 🛛 🛛	município ×
ocalidade i ×		
otas i ×		

Assim como o bloco *identificação*, o bloco *coleta* oferece a busca em campos específicos referentes à coleta da amostra. Muitos registros descrevem o local da coleta e acrescentam as notas do coletor. Assim como o campo do determinador, é comum encontrar o nome do coletor com grafias distintas.

#### **Bloco coleção**

coleção 1			
tipo		rede	
	×		× 1
acrônimo 🗙 🕨			
núm. catálogo 🛛 🕹		cód, barras 🛛 👋	

Este bloco permite buscas pelo tipo de coleção (botânica, microbiológica, zoológica e abrangente); por rede (INCT-HVFF, SICol, Polinizadores, Taxonline e iDigBio); pelo acrônimo da coleção, número de catálogo do espécime, ou código de barras.

#### **Bloco microbial**

Esse bloco é específico para coleções microbianas, permitindo a busca pelo histórico de depósito, risco biológico, propriedades e hospedeiro ou substrato.

### Bloco filtros geográficos

Os filtros geográficos funcionam como máscaras geográficas que restringem a

microbial i	
histórico de depósito 🛛 🛛	
risco biológico 🛛 ×	
propriedades ×	
hospedeiro ou substrato	×

busca a apenas registros que caiam dentro dela. Assim, ao selecionar, por exemplo, o bioma



Pampa como filtro, apenas os registros cujas coordenadas caiam dentro do polígono que define o bioma serão utilizados. Esse filtro pode ser combinado com todos os outros textuais.

O sistema permite o uso simultâneo de mais de um filtro geográfico. É possível, por exemplo, recuperar os registros do Bioma Cerrado do Estado de Minas Gerais. No mapa gerado pelo sistema, observa-se a intersecção do Bioma Cerrado com o Estado de Minas Gerais e o ponto de ocorrência de um espécime.

#### Bloco Uso e cobertura da terra MapBiomas



#### MAPBIOMAS

💻 uso e cobertura da terra - MapBiomas 🕨

Anualmente o MapBiomas realiza o mapeamento da cobertura e uso da terra para todo o Brasil e países sul americanos. São dados disponibilizados a partir de 1985 de forma livre, aberta e *online*. Cada registro no *species*Link, coletado no Brasil ou na Panamazônia, com coordenadas originais, tem os dados de uso e cobertura da terra do MapBiomas a ele associados. Até mesmo as coletas realizadas em anos anteriores a 1985 têm a sua coordenada indexada para esse período.

#### Bloco Dados agregados e modificadores

Além dos dados do MapBiomas, outras informações complementares como a qualificação taxonômica, forma de vida, endemismo e origem das espécies botânicas, extraídas da flora e funga do Brasil, grau de ameaça e qualidade e origem das coordenadas geográficas também são agregadas aos registros.

Os filtros para esses dados compõem o último conjunto da interface de busca. O uso desses filtros possibilita a inclusão dos sinônimos de uma espécie na busca além de filtrar a busca de acordo com a base do registro, imagens, material tipo, forma de vida, endêmico do Brasil e origem, de acordo com a Flora e Funga do Brasil, espécies ameaçadas de extinção (todas as categorias), status taxonômico e qualificação das coordenadas geográficas.

<ul> <li>incluir sinonímia i</li> <li>AlgaeBase</li> <li>MycoBank</li> <li>Flora e Funga do Brasil</li> <li>Catálogo Moure</li> <li>DSMZ</li> <li>GBIF</li> </ul>	base do registro     espécime preservado     espécime vivo     espécime fóssil     observação humana     observação por instrumento     amostra de material     com anotações	imagens     com imagens     com imagens     material vivo     pólen     madeira     microscopia     sem imagens	<ul> <li>material tipo</li> <li>só typus</li> <li>nome espécie tipo</li> <li>na flora e funga do brasil i forma de vida</li> <li>endêmico brasil</li> <li>origem</li> </ul>
<ul> <li>espécies ameaçadas (brasil)</li> <li>CR - Criticamente em Perigi</li> <li>CR (PEX) - Provavelmente E</li> <li>EN - Em Perigo</li> <li>EW - Extinta na Natureza</li> <li>EX - Extinta</li> <li>RE - Extinta no Brasil</li> <li>VU - Vulnerável</li> </ul>	i IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	status taxonômico <i>i</i> aceito com autor aceito sem autor* sinônimo com autor sinônimo sem autor* ambíguo sem autor* ambíguo com autor não encontrado só gênero	<ul> <li>coordenadas geográficas</li> <li>por município</li> <li>originais</li> <li>consistente</li> <li>suspeitas</li> <li>bloqueadas</li> </ul>

### Formas de visualizar o resultado de uma busca (outputs)

A rede oferece várias formas de visualizar o resultado de uma busca. As opções ou *outputs* são: REGISTROS, INVENTÁRIOS, MAPA, IMAGENS, NÚMEROS e DUPLICATAS

No exemplo dos outputs vamos utilizar os seguintes filtros de busca expressos na figura a seguir.



Assim, a busca que servirá de exemplo usará os dados da rede **INCT-HVFF**, buscando registros de **árvores nativas**, **endêmicas** do Brasil, que são **espécies ameaçadas**, com status taxonômico **aceito com autor** e com **coordenadas originais** (informadas pela coleção) e **consistentes**. Essa busca em 04/10/2024 recuperou 14.973 registros que serão utilizados na visualização dos resultados da busca.

#### REGISTROS

Efiltros Eregistros Inventários Amapa Olimagens Números Puplicatas Amaxia Como citar

Podemos observar que a opção selecionada **REGISTROS** agora é apresentada no cabeçalho em laranja. A seguir é apresentado um dos 14.973 registros recuperados nessa busca.

ACTION OF COMPANY OF COMPANY.	PLANTAE RUBIACEAE Melanopsidium nigrum Colla. Det: Valadares, R.T. 16/1/2020 ameaçada - VU VIES 46201 Coleta: Firmino, A.D.; Matos, N.G.; Penha, J.S.; Pereira, O.J. 1468 04/12/2019. Loc: Aldeia dos Cocos, ponto de amostragem 3., São Mateus, Espírito Santo, Brasil Cód. barras: VIES046201. Coord. orig.: [/at: -18.7869444 <i>long</i> : -39.7505556 WGS84] ⊡ Notas: Tronco com casca esbranquiçada. Flores alvas, internamente amarelo claro Habitat: Arbórea Habito: Árvo Base do registro: PreservedSpecimen	re	*
	Status taxonômico: aceito, Flora e Funga do Brasil <sup>[2]</sup> .     Nome encontrado com autor		
	• Classificação GBIF •• Plantae » Tracheophyta » Magnoliopsida » Gentianales » Rubiaceae » <i>Melanopsidium nigrum</i> Colla 🗗 accepted		
	• Na Flora e Funga do Brasil •• Endemismo: É endêmica do Brasil• Origem: nativa• Forma de vida: Arbusto; Árvore		
	Uso e cobertura da terra - MapBiomas Brasil Col. 8 (1985-2022)		
9	VIES-Plantas - Herbário Central da Universidade Federal do Espírito Santo VIES 🔽	nova anotação	detalhes

Podemos visualizar o trabalho de qualificação de determinados dados do registro como: nome científico (negritado e de cor verde que indica ser um nome aceito com autor), o *tag* de espécie ameaçada VU e o *link* da coordenada da coleta que mostrará a sua localização em um mapa, o *link* para o status taxonômico na referência Flora e Funga do Brasil, o *link* para a classificação GBIF e o *link* para os dados cadastrais do Herbário que mantém essa exsicata.

Passando o mouse no esquema do uso e cobertura da terra, podemos observar que a coordenada onde foi realizada a coleta, entre 1985 a 1990 é definido como *Mosaico de Usos* que, a partir de 1991 é classificado como *Restinga Arborizada*, inclusive no ano da coleta da espécie em 2019.

### **INVENTÁRIOS**

EFILTROS REGISTROS INVENTÁRIOS AMAPA O IMAGENS NÚMEROS D'UPLICATAS AMAXAR COMO CITAR

Sempre utilizando os mesmos parâmetros de busca, podemos visualizar o resultado como **INVENTÁRIO**. Os inventários disponíveis são: coleção, base dos registros, categoria de ameaça, ano da coleta, mês da coleta, país, estado, ano da determinação, *typus*, coletor, reino, filo, classe, ordem, família, gênero, nome da espécie, status taxonômico, município e coletor. Usaremos como exemplo o "mês da coleta".



mês da coleta 👻		
01	1194	10.1%
02	1142	9.7%
11	1117	9.5%
03	1094	9.3%
10	1094	9.3%
09	1031	8.7%
04	954	8.1%
05	937	7.9%
07	898	7.6%
08	810	6.9%
12	806	6.8%
06	728	6.2%
13 itens	11805	
em branco	468	
Lxcel	⊥ TS	V

O usuário pode ordenar o inventário de acordo com o mês da coleta ou número de registros. Vários tipos de gráficos estão disponíveis.

#### MAPA

🗑 FILTROS 🗊 REGISTROS 🗎 INVENTÁRIOS 🗣 MAPA 🔯 IMAGENS 📓 NÚMEROS 🖓 DUPLICATAS 🕹 BAIXAR 📓 COMO CITAR

*species*link

A opção mapa apresenta os registros recuperados em um mapa. Os layers de fundo disponíveis online incluem *Base, Topografia, Satélite, StreetMap* e o *MapBiomas*. A figura a seguir apresenta os registros plotados no mapa base.



#### **IMAGENS**

BFILTROS ■REGISTROS ■INVENTÁRIOS ♥MAPA <sup>®</sup>IMAGENS <sup>®</sup>NÚMEROS <sup>®</sup>DUPLICATAS <sup>®</sup>BAIXAR <sup>®</sup>COMO CITAR <sup>®</sup>Clicando na opção IMAGENS, o usuário receberá uma ou mais páginas com as imagens recuperadas em formato *thumbnail*. A seguir apresentamos algumas imagens em uma das páginas e uma imagem em destaque. Ao clicar na imagem pequena, ela será aberta em um visualizador que permite fazer zoom e outras funcionalidades.







O resultado de uma busca também pode ser expresso em números. Lembrando que continuamos com a busca de registros da rede INCT-HVFF de árvores nativas, endêmicas, ameaçadas de extinção com status taxonômico aceito com autor e coordenadas originais consistentes. Essa busca expressa em NÚMEROS é:

161 conj. dados	137 instituições			<b>14,973</b> registros <i>i</i>
imagens				
7,836 reg. com imagens	7,137 reg. sem imagens			<b>9,676</b> imagens <i>i</i>
nomes científicos	s distintos		registros identifi	cados
1,577 aceitos	0 sinônimos		14,973 até espécie	0 até gênero
<b>O</b> ambíguos	0 não encontrados		0 até família	0 sem identificação
coordenadas geo	gráficas			
14,973 originais	0 por município	<mark>0</mark> bloqueadas		14,973 reg. georref.
duplicatas <mark>i</mark>				
<b>19,616</b> regs analisados <u>i</u>	2,369 dups conjunto <i>i</i>	6,239 dups rede <mark>i</mark>		11,008 coletas distintas <u>i</u>
outros números				
167 reg. com anotações	1,373 reg. de <i>typus</i>	14,973 reg. espécies ameaçadas		

Os registros recuperados que atenderam os critérios de busca utilizados, têm como fonte 161 conjuntos de dados de 137 instituições. Foram recuperados 14.973 registros, sendo 7.836 com imagens, totalizando 9.676 imagens. São 1.577 nomes científicos distintos aceitos e 14.973 registros identificados até espécie, todos com coordenadas originais.

Foram analisados 19.616 registros para avaliar as duplicatas tanto no conjunto dos dados da busca, bem como em toda rede *species*Link. Foram encontradas 2.369 duplicatas no conjunto analisado e mais 6.239 duplicatas em toda rede. Os registros provêm de 11.008 coletas distintas. O sistema apresentou 167 registros com anotações, 1.373 registros de *typus* e 14.973 registros de espécies ameaçadas de extinção.

Esse é apenas um exemplo. O resultado de qualquer busca no *species*Link pode ser expresso em números.



O usuário poderá também verificar as duplicatas do conjunto de dados de sua busca. A figura a seguir mostra um registro com quatro duplicatas.





### BAIXAR

FILTROS PREGISTROS INVENTÁRIOS AMAPA O IMAGENS NÚMEROS CO DUPLICATAS BAIXAR COMO CITAR

Para **baixar** (download) o resultado de uma busca, o usuário precisa estar cadastrado no *species*Link. O cadastro é necessário porque o link que permite baixar o resultado da busca será enviado por email. O cadastramento é gratuito, exige apenas o nome e email do usuário e dá acesso a funcionalidades extras do *species*Link.

### COMO CITAR

🖺 FILTROS 🗊 REGISTROS 🖺 INVENTÁRIOS 🕈 MAPA 🔯 IMAGENS 📓 NÚMEROS 🖓 DUPLICATAS 🕹 BAIXAR 📓 COMO CITAR

Por fim, a opção **COMO CITAR** apresenta os filtros utilizados na busca, o número total de registros recuperados, os conjuntos (coleções) que forneceram dados para essa busca e número de registros oferecidos por cada coleção.

A interface de busca permite diferentes análises. Ao clicar no componente MapBiomas o usuário visualiza duas colunas com as mesmas opções de busca, podendo selecionar o **ano da coleta**, ou qualquer ano a partir de 1985, e especificar os tipos de cobertura ou uso da terra no local onde o espécime foi coletado para restringir os registros a serem considerados na busca.

*species*link



O usuário pode também buscar o "antes" e "depois" para observar as transformações ocorridas no ponto da coleta.

No exemplo a seguir, foram utilizados os seguintes critérios de busca: tipo de coleção: **botânica**; base do registro = **espécime preservado**, forma de vida = **árvore**; origem = **nativa**, status taxonômico = **aceito com autor**, com coordenadas geográficas = **originais e consistentes**, coletas no estado = **Pará**, em **áreas naturais (MapBiomas)**.

O resultado expresso em NÚMEROS é:



São mais de 12 mil registros de árvores nativas com coordenadas originais consistentes, coletadas no estado do Pará em áreas naturais.

Podemos incluir um novo elemento nessa busca em relação aos dados do MapBiomas como registros coletados em áreas naturais, que em 2023 são áreas antrópicas. Assim temos como elementos de busca: coleção: **botânica**; base do registro = **espécime preservado**, forma de vida = **árvore**; origem = **nativa**, status taxonômico = **aceito com autor**, com coordenadas geográficas = **originais e consistentes**, coletas no estado da coleta = **Pará** em **áreas naturais, que em 2023** são áreas antrópicas.

uso	eco	bertura da terra em	uso e cobertura da terra em
a	io di	a coleta 🗸 🧹 Natural 🗌 Antrópico	2023 🗸 🗌 Natural 🗹 Antrópic
<b>~</b>	Flo	esta	Floresta
		Formação florestal.	🗖 Formação florestal.
		Formação savânica	Formação savânica
		Mangue	Mangue
		Floresta alagável	Floresta alagável
		Restinga Arborizada	Restinga Arborizada
	For	mação natural não florestal	Formação natural não florestal
		Campo alagado e área pantanosa	Campo alagado e área pantanosa
		Formação campestre	Formação campestre
		Apicum	Apicum
		Afloramento rochoso	Afloramento rochoso
		Restinga Herbácea	🗌 Restinga Herbácea
		Outras Formações não Florestais	Outras Formações não Florestais
	Agr	opecuária	Agropecuária
		Pastagem.	Pastagem.
		Agricultura	Agricultura
		🗌 Lavoura temporária	Lavoura temporária
		🗋 Soja	🗹 Soja
		Cana Cana	Cana Cana
		Arroz	Arroz
		Algodão	Z Algodão
		Outras Lavouras Temporárias	Outras Lavouras Temporárias
		Lavoura perene	Lavoura perene
		Café Café	Café Café
		Citrus	Citrus
		Dendê	Z Dendê
		Outras Lavouras Perenes	Outras Lavouras Perenes
		Silvicultura	Silvicultura
		Mosaico de Usos	Mosaico de Usos
	Áre	a não vegetada	🗹 Área não vegetada
		Praia, duna e areal	Praia, duna e areal
		Área Urbanizada	🗹 Área Urbanizada
		Mineração	Mineração
		Outras Áreas não Vegetadas	💙 Outras Áreas não Vegetadas
	Cor	po dágua	Corpo dágua
		Rio, Lago e Oceano	🔲 Rio, Lago e Oceano
		Aquicultura	Z Aquicultura
		Geleira	🖸 Geleira
	Não	observado	🗹 Não observado

O resultado da busca apresentado em NÚMEROS é:

25 conj. dados	22 instituições			244 registros i
imagens				
160 reg. com imagens	84 reg. sem imagens			<b>219</b> imagens <i>i</i>
nomes científicos	distintos		registros identifi	cados
150 aceitos	0 sinônimos		244 até espécie	0 até gênero
<b>O</b> ambíguos	0 não encontrados		0 até família	0 sem identificação
coordenadas geo	gráficas			
244 originais	0 por município	0 bloqueadas		244 reg. georref.
duplicatas <mark>i</mark>				
317 regs analisados <mark>i</mark>	30 dups conjunto <i>i</i>	95 dups rede i		192 coletas distintas <mark>i</mark>
outros números				
2 reg. com anotações	0 reg. de typus	1 reg. espécies ameaçadas		

É possível produzir um inventário do uso e cobertura da terra das espécies no ano da coleta e compará-lo com a situação em 2023.



Os especialistas podem, por exemplo, produzir uma lista das espécies que foram coletadas em áreas naturais que em 2023 são pastagens. Avaliar as espécies disponíveis em viveiros e recomendá-las para uso na restauração. O gráfico de nomes de espécies a seguir, apresenta essas espécies cujo local de coleta em área natural se tornou pastagem no estado do Pará.



*species*link



É evidente que a análise para indicar espécies para restauração é muito mais complexa e deve ser feita por um especialista. Esse documento pretende mostrar os recursos disponíveis na interface que eventualmente podem auxiliar nesse trabalho.

### A importância dos filtros de busca

O objetivo dos dos filtros é viabilizar a recuperação dos registros que atendem os requisitos do usuário. A princípio, essa interface pode parecer complexa, mas uma vez que o usuário entende quais são as opções de busca e as opções de visualização do resultado da busca, o processo se torna mais simples e as diferentes possibilidades de análise muito interessantes.

Ao analisar o resultado de uma busca, é comum querer excluir ou incluir novos elementos na busca. Para tanto, só é necessário selecionar a opção FILTROS novamente para incluir ou excluir elementos da busca e, em seguida, selecionar a opção desejada para visualizar o resultado.

Ao compreender esse procedimento, a busca se torna mais simples e, com as diferentes formas de visualizar o resultado da busca apresentadas a seguir, o sistema possibilita uma análise mais ampla dos dados recuperados.

### Data Cleaning

Um componente importante do *species*Link são as ferramentas para avaliar a qualidade dos dados, destacando possíveis erros. Desde o início do desenvolvimento da rede *species*Link haviam conceitos muito claros:

- 1. O curador ou responsável pela coleção teria total domínio sobre seus dados; e,
- 2. A equipe do CRIA não faria qualquer alteração nos dados.

No entanto, verificamos que existem erros simples, mas de difícil identificação, dado o número de registros na coleção. Foi então desenvolvido um conjunto de aplicativos para ressaltar esses **possíveis** erros, cabendo sempre ao curador a análise do relatório *dataCleaning*, corrigir possíveis erros e reenviar os registros para a rede.

A ferramenta *dataCleaning* precisa ser reescrita, sendo que esse processo já teve início. Apesar da interface antiga (*splink.cria.org.br/dc*) estar desatualizada, particularmente no que se refere à checagem de nomes científicos, continua disponível, mas deve ser usada com cautela.

Muitos dos possíveis erros hoje podem ser detectados a partir da interface de busca.

Utilizaremos como exemplo o herbário CAS-BOT da *California Academy of Sciences* que compartilha 22.712 registros com a rede *species*Link..

O primeiro bloco do dataCleaning apresenta os seguintes dados gerais:

coleção: CAS-BOT				
total de registros indexados	22712			
- sem coordenadas geográficas	11755			
- georreferenciados	10957			
- georreferenciados bloqueados	0			
- no mar	137			
- número do catálogo em branco	0			
menor: 4 maior: 726576	[ gap ]			
registros repetidos				
número do catálogo	0			
registros duplicados	0			
nome + número do coletor	1419			
última atualização - logs de erro				
da coleção: 20-08-2024 do dataCleaning:	20-09-2024			

É importante entender que o relatório indica possíveis erros. O quadro, por exemplo, indica 137 registros de ocorrência no mar. Se o herbário não tem nenhuma coleta no mar é só clicar nesse número para receber uma lista desses registros para avaliação. Se essas coletas foram realizadas no mar, é só ignorar. Outros componentes do *data cleaning* incluem data de coleta, dados geográficos e sugestões do sistema para campos em branco. Nesse momento enfaticamente desencorajamos o uso dos relatórios que tratam de nomes científicos por estarem desatualizados.

### Interface de busca como ferramenta para encontrar possíveis erros

É importante analisar o uso da interface de busca para encontrar possíveis erros e inconsistências. Os filtros existentes permitem a busca pelo status taxonômico dos registros (aceito, sinônimo, ambíguo com ou sem autor ou não encontrado) e as coordenadas geográficas originais, consistentes, suspeitas e o georreferenciamento por município. Assim o curador da coleção poderá analisar os dados suspeitos, enquanto o usuário poderá filtrar sua busca por nomes aceitos com coordenadas consistentes.

A rede *species*Link utiliza as seguintes referências para qualificar os nomes: AlgaeBase, MycoBank, Flora e Funga do Brasil, Catálogo Moure, DSMZ, e GBIF. Cada nome científico é indexado de acordo com o seu status taxonômico: aceito com autor, aceito sem autor, sinônimo com autor, sinônimo sem autor, ambíguo sem autor, ambíguo com autor, não encontrado e só gênero.

Expressando o resultado da busca por todo acervo do herbário CAS-BOT em números temos uma primeira ideia sobre a qualidade dos nomes científicos:



A interface de busca permite identificar todos os registros com nomes que são aceitos, sinônimos, ambíguos ou não encontrados.

### Estatísticas de uso dos dados

### specieslink.net/usage

Essa ferramenta foi desenvolvida para que as redes e coleções que compartilham seus dados com a rede *species*Link possam apresentar um indicador de acesso/uso dos dados. A ferramenta apresenta os seguintes filtros:

### Estatísticas de uso dos dados

Rede	Тіро	Coleção	Dia	Mês	Ano	
~	~	~	~	~	~	Exibir

O usuário poderá visualizar as estatísticas de uso de determinadas redes, tipos de coleção, coleções específicas e ainda produzir estatística para um determinado dia, mês, ou ano. Uma busca pela rede INCT Herbário Virtual desde a implementação desse indicador, no dia 17/10/2024, gerou os gráficos dos registros e imagens utilizadas apresentados a seguir.

#### INCT Herbário Virtual



### **Registros utilizados**



As tabelas e gráficos apresentam o número de registros da rede INCT-HVFF utilizados do dia 10 de outubro de 2012 ao dia 14 de outubro de 2024. Em 2019 e 2020, o número de registros utilizados por ano era da ordem de 530 milhões para um acervo online de 10,6 milhões de registros. Em 2021 a rede teve significativos avanços, particularmente em relação à tecnologia utilizada.

Graças aos avanços tecnológicos, em 2023, o número de registros utilizados saltou para 62 bilhões. Atualmente o INCT-HVFF compartilha 13,3 milhões de registros online (outubro de 2024). 62 bilhões de registros utilizados em 2023 representam uma média de 170 milhões de registros botânicos e micológicos utilizados por dia através da interface de busca da rede *species*Link. Como o INCT-HVFF no dia desta análise (14/10/2024) oferece 13,3 milhões de registros, é como se os usuários utilizassem diariamente cerca de 13 vezes todo o acervo online.

Essa ferramenta é importante para todos os provedores de dados da rede *species*Link. Cada coleção biológica pode acessar as suas estatísticas e assim mostrar a importância e alcance do seu acervo.

O sistema também apresenta as estatísticas das imagens visualizadas.

### Imagens utilizadas



Toda preparação e serviço das imagens online foi objeto de um novo desenvolvimento. Em 2021 o número de imagens utilizadas saltou de 3 para 6 milhões ao ano. Em 2022, com a nova interface estabelecida, esse número saltou para cerca de 18 milhões de imagens utilizadas por ano. A partir de 2023 esse número é maior de 32 milhões de imagens utilizadas por ano.

O INCT-HVFF hoje (14/10/2024) oferece mais de 6,2 milhões de imagens associadas a 5,4 milhões de registros. Novamente usando os dados do ano de 2023, temos uma média de 88,7 mil imagens visualizadas por dia, sendo que somente são computadas as imagens visualizadas através da ferramenta *mosaico*. As imagens visualizadas na lista inicial dos registros recuperados na busca, quer em tamanho pequeno, quer em tamanho maior, não são computadas.

As imagens antes organizadas e oferecidas online através de um software proprietário, também foram reprocessadas usando software aberto para serem disponibilizadas no Google Cloud Platform a partir de 2022.

### Dashboard

### specieslink.net/dashboard

Outro indicador da rede *species*Link é o *Dashboard*, cujos dados podem ser analisados para toda rede *species*Link, as diferentes redes ou por tipos de coleções. Os dados do INCT-HVFF no dia 17/10/2024 são apresentados nas figuras a seguir.

### Dashboard

Rede	Tipo				
INCT Herbário Virtual	~	✓ filtr	ar		
Conteúdo geral		Classificação dos regi	stros	Provedores de dados	
Conjuntos de dados	247	Por grupo taxonômico		Empresa Privada	1
Total de registros	13.327.912	Botânica	13.327.912	Empresa Pública	4
Com Imagens	5.468.476	Microbiologia	1.820	Instituição do Exterior	
Georreferenciamento		Por tipo de registro		Instituto Estadual	11
Registros georreferenciados	9.618.860	Fóssil	5.744	Instituto Federal	5
Registros georreferenciados por	2,993,364	Observação humana	0	Instituto Federal Educação e C&T	1
município	2.550.001	Espécime vivo	0	Instituto Municipal	2
Registros com coordenadas	6.625.496	Observação por máquina	14.278	ONG/OS	2
Espécies		Espécime preservado	13.295.417	Secretaria Estadual	- 0
Registros de espécies ameaçadas	129.717	Outros	10.844	Universidade Estadual	22
Registros com nomes aceitos	9.237.244				20
Nomes aceitos distintos	216.067			Universidade Federal	53
Registros de typus	365.820			Universidade Municipal	1
				Universidade Particular	11
				Pessoa física	1
				Total	142

Nessa parte do dashboard temos dados gerais sobre o conteúdo. Um dado interessante que é recuperado através da interface de busca é a qualificação dos provedores de dados, predominantemente das universidades e Institutos de pesquisa.



Registros por país (América		Registros por UF (Brasil)			
do Sul)		Acre	82.398		
Argentina Bolívia Brasil Chile Colômbia Equador Guiana Francesa Paraguai Peru Suriname Uruguai Venezuela <b>Reguistros por ri</b>	317.521 792.510 7.859.049 163.181 644.645 915.938 206.880 132.869 263.255 1.000.309 111.362 35.871 490.454	Alagoas Amazonas Bahia Ceará Distrito Federal Espírito Santo Golás Maranhão Mato Grosso Mato Grosso Mato Grosso Mato Grosso do Sul Minas Gerais Pará	81.546 59.372 429.795 863.286 107.187 118.475 235.170 244.08 95.202 256.142 198.157 915.995 307.165 84.124		
Centro Oeste Norte Nordeste Sul Sudeste	948.838 1.044.230 1.618.772 1.483.601 2.177.914	Parambuco Plauí Rio de Janeiro Rio Grande do Norte Rio Grande do Sul Rondônia Roraima Santa Catarina São Paulo Sergipe Tocantins	610.708 234,355 51.522 32.0662 98.213 67.287 343.765 706.087 62.954 44.456		

O segundo bloco mostra a localização geográfica das coleções que compartilham seus dados com o INCT-HVFF. Além dos herbários brasileiros, o INCT-HVFF estabeleceu parcerias com herbários dos Estados Unidos e Europa. Essa parceria inicialmente visava somente a integração de dados sobre coletas realizadas no Brasil. O escopo geográfico foi ampliado para toda a América do Sul, como mostra o Dashboard. Com exceção da Caatinga, todos os demais biomas brasileiros extrapolam nossa fronteira política. Por fim, o terceiro bloco mostra a evolução da quantidade dos dados e imagens online.



### API do serviço web v1.0

specieslink.net/ws/1.0/

O objetivo da API é facilitar o acesso aos dados da rede *species*Link por outros programas e aplicativos. As chamadas devem ser feitas através do método HTTP GET, com possibilidade de passar parâmetros como parte da *query string*. As respostas são sempre em formato JSON. Este serviço oferece acesso tanto aos metadados dos provedores, como nome e tipo das coleções, quanto aos registros de biodiversidade. O serviço web é aberto a todos, porém requer uma chave de acesso que deve ser fornecida a cada chamada ao serviço através do parâmetro "apikey". Acesse a API para uma explicação mais detalhada do serviço.

### Cadastro do usuário

O acesso *online* aos dados da rede *species*Link é livre e aberto a todos os interessados, mas existem algumas ferramentas que necessitam que o usuário seja cadastrado. São elas: (1) para fazer o download dos dados, uma vez que o link para baixar o arquivo com o resultado será enviado por email; (2) o mesmo se aplica à ferramenta *Anotação*, onde o usuário pode sugerir uma correção ou até uma nova determinação do espécime; e (3) para usar a interface de programação de aplicações (API). A página de busca do *species*Link apresenta em seu cabeçalho um breve texto sobre a necessidade do cadastro e o acesso ao formulário. Para se cadastrar basta clicar na opção **Entrar**.



Além de cadastrar uma senha, são solicitados ao usuário dados como nome, email e instituição. São também solicitadas as principais motivações para uso do *species*Link. São cerca de 15 mil usuários cadastrados, sendo que 98% indicaram suas motivações para uso da rede. Trata-se de uma informação importante para assegurar que a rede continua sendo utilizada pelo seu público alvo.



### Motivações para uso do sistema

(mais de uma motivação pode ser marcada pelo mesmo usuário)

pesquisa (90%)
educação (55%)
conservação (33%)
consultoria ambiental (21%)
curadoria de coleção (13%)
divulgação científica (13%)
restauração ecológica (13%)
EIA/RIMA (12%)
política pública (7%)
bioeconomia (5%)
outro (5%)