

Разнообразие почвенных животных России: публикация и эффективное использование исходных данных  
29–30 августа 2019 г., ИПЭЭ им. А.Н. Северцова РАН

Сессия 8.

# Как опубликовать данные в GBIF

## Что такое IPT?

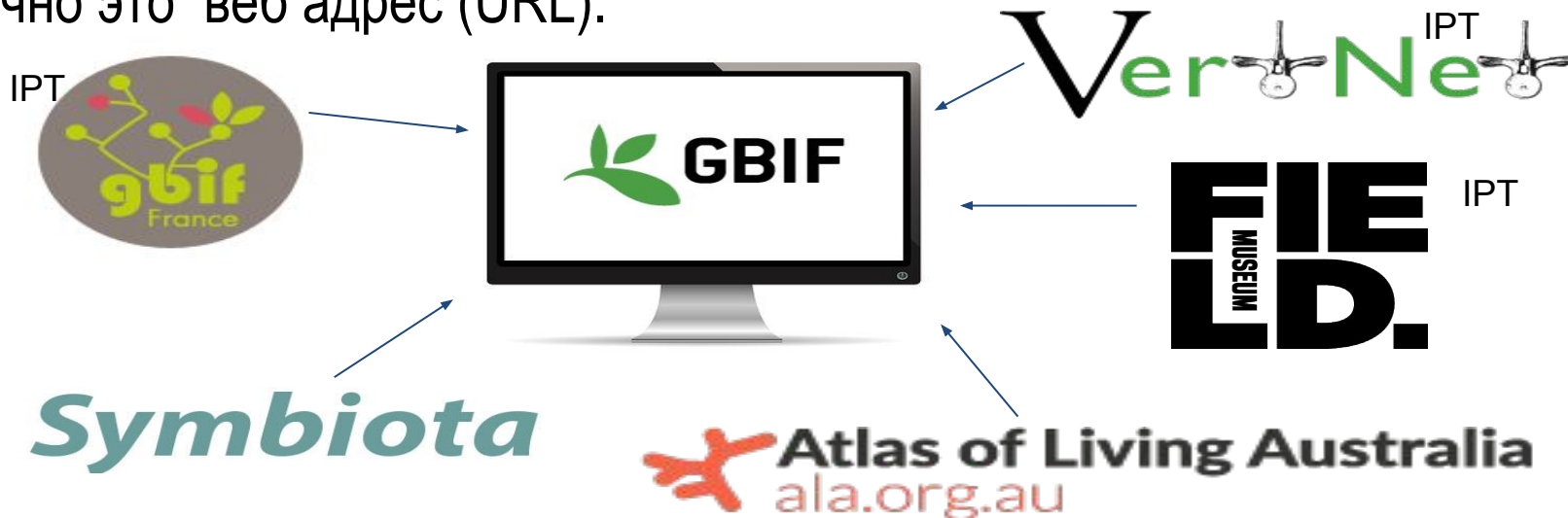
**Максим Шашков**

Институт математических проблем биологии РАН – филиал ИПМ им. М.В. Келдыша РАН  
г. Пущино Московской области

**Слайды CC BY:**  
*Marie Elise Lecoq,  
Sophie Pamerlon,  
Nicolas Noé и  
Максим Шашков*

# Публикация

“Публикация” - способ сделать **наборы данных** о биоразнообразии **общедоступными** и **доступными** для обнаружения в **стандартизированной форме** через **точку доступа**,  
обычно это **веб адрес (URL)**.



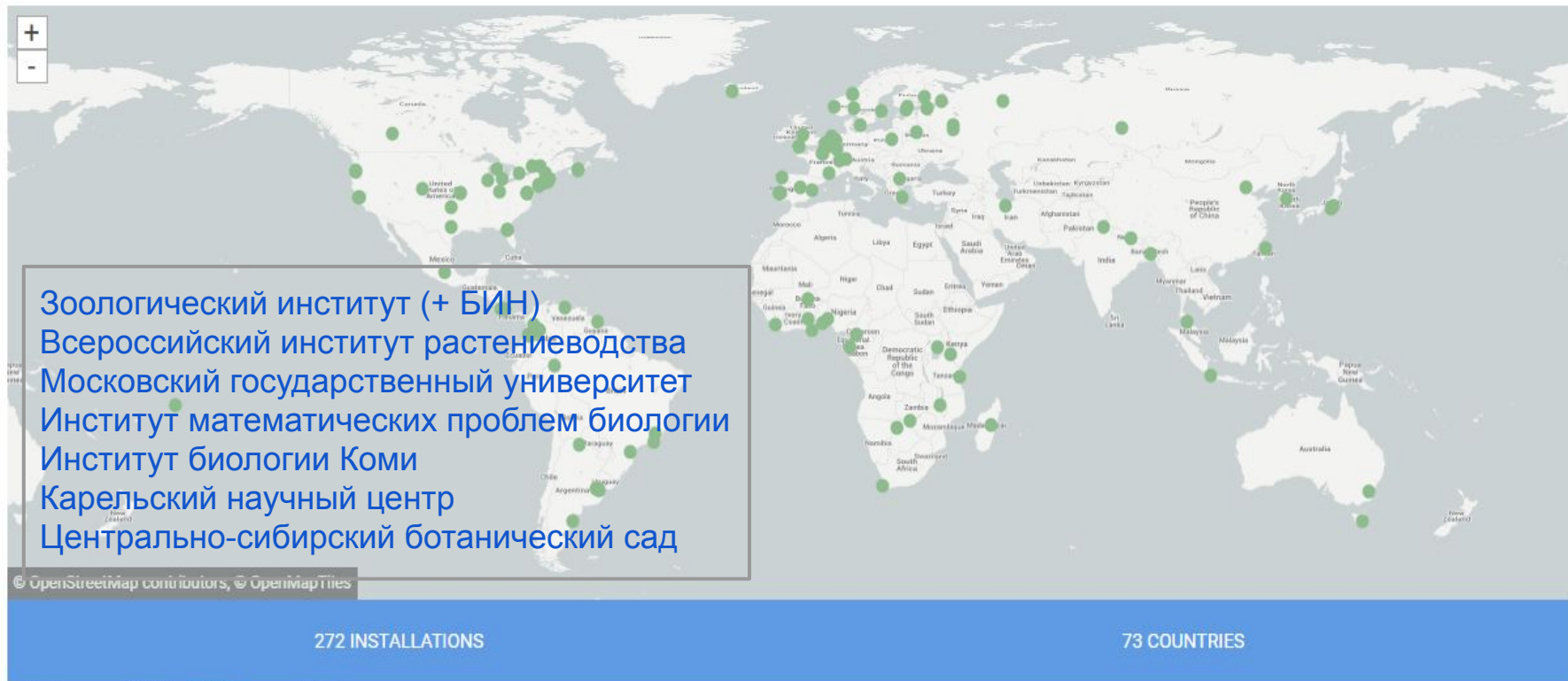
# Метод публикации данных: IPT

- **Серверное программное обеспечение** для работы необходимо стабильное подключение к сети Интернет
- Один IPT может поддерживать несколько наборов данных, опубликованных разными организациями, с сохранением авторства
- **Основной** (но не единственный) инструмент для публикации в GBIF
- IPT может работать в **тестовом** и **рабочем режиме**
- IPT поддерживает **многоязычность**

# Сеть IPT

## IPT: The Integrated Publishing Toolkit

*A free open source software tool used to publish and share biodiversity datasets through the GBIF network.*

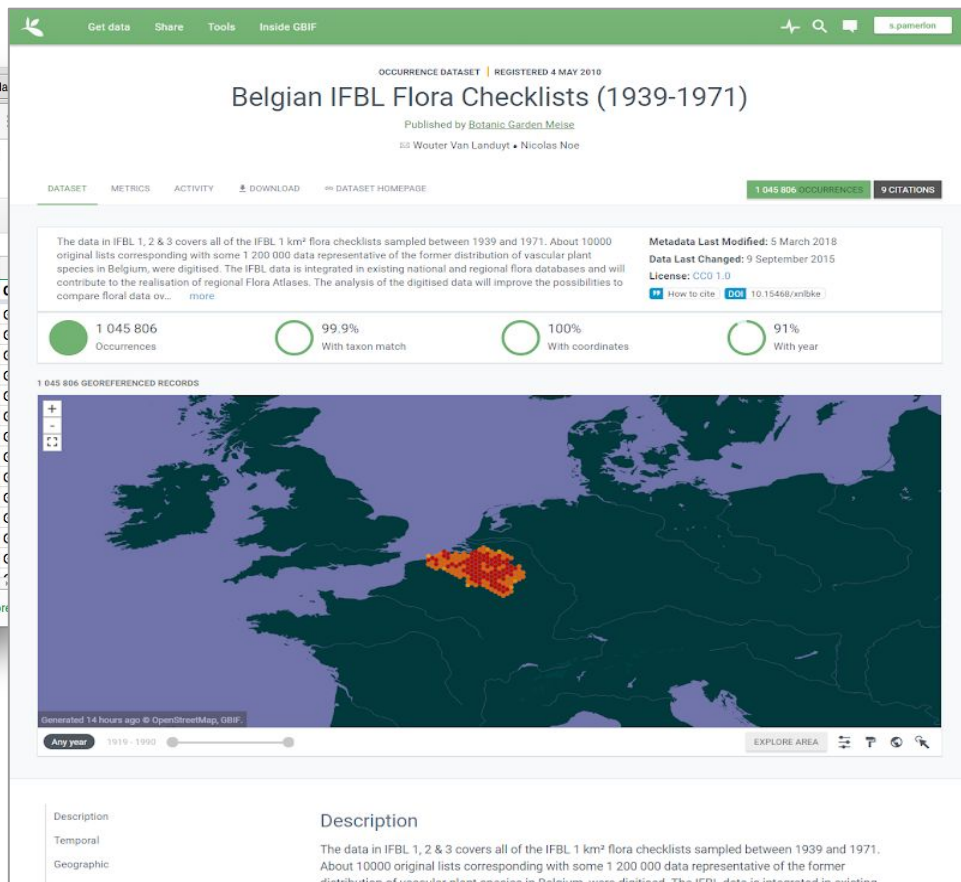


Don't see your IPT? Send [GBIF](#) your coordinates.

# Публикация

BID CE Workshop 1 - SESSION 08 - DATA EXAMPLE

Cat. Numb.	University	Collector	No. of spec.	YE	MO	DA
1	UWP:100217	University of Guatemala	Betancur J	1	1991	5
2	UWP:100218	University of Guatemala	Betancur J	1	1991	5
3	UWP:101378	University of Guatemala	Fonnegra R	1	1994	5
4	UWP:101717	University of Guatemala	Betancur J	1	1993	11
5	UWP:101737	University of Guatemala	Betancur J	1	1993	11
6	UWP:102143	University of Guatemala	Betancur J	1	1994	4
7	UWP:102144	University of Guatemala	Betancur J	1	1994	4
8	UWP:102233	University of Guatemala	Vargas I	1	1995	12
9	UWP:103108	University of Guatemala	Cardona F	1	1996	4
10	UWP:104139	University of Guatemala	Fonnegra R	1	1996	5
11	UWP:104512	University of Guatemala	Callejas R	1	1995	7
12	UWP:105292	University of Guatemala	Acevedo P	1	1994	9
13	UWP:106768	University of Guatemala	Vargas I	1	1997	4

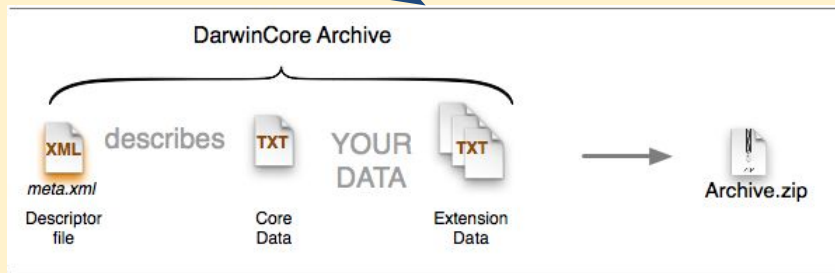
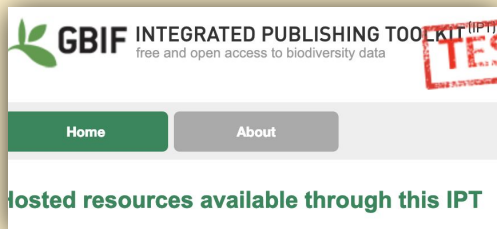


# IPT - технические подробности



= Программное обеспечение

Работает под управлением веб-сервера (Tomcat)



IPT создает DwC архив,  
который затем становится  
доступным через сеть GBIF

# IPT - этапы процесса публикации

1. Создать новый набор данных
2. Загрузить таблицу
3. Сопоставить поля с терминами (две концепции: заголовки с вашими именами, а затем сопоставляете, либо имена полей по DwC; для полей с единственным значением для всех записей - константы можно задать при сопоставлении)
4. Заполнить метаданные
5. Опубликовать
6. Зарегистрировать



# IPT - ядра Darwin Core

На этапе сопоставления полей таблицы необходимо выбрать “ЯДРО” стандарта Darwin Core соответствующее подготовленным данным

**Occurrence Core** для находок

**Taxon Core** для списков видов

**Event Core** для описаний, учётов и сборов

**Ядро (Core)** - это набор терминов Darwin Core для описания определенного типа данных, в каждом ядре есть свой перечень обязательных полей

The screenshot displays the GBIF Integrated Publishing Toolkit (IPT) interface. At the top, the GBIF logo is followed by the text 'INTEGRATED PUBLISHING TOOLKIT (IPT)' and 'free and open access to biodiversity data'. Below this is a navigation bar with 'Home', 'Manage Resources' (highlighted in green), and 'Administration' buttons. The main content area shows the 'Overview' page for the 'carabus\_test' resource. It includes a section for 'Source Data' with a file upload button, a 'Connect to database' button, and a 'Clear' button. Below this is a 'Darwin Core Mappings' section with a dropdown menu showing 'Darwin Core Occurrence' selected, and a list of other mappings: 'Core', 'Darwin Core Occurrence', 'Darwin Core Taxon', and 'Darwin Core Event'. The interface also shows a warning icon and text indicating that the mapping is between the source data and the resource metadata.



# Тип данных: список видов (Checklist)

	A	B	C	D	E	F	
1	taxonID	parentNameUsageID	parentNameUsage	acceptedNameUsageID	acceptedNameUsage	scientificName	nameAccording
2	73			73	Equisetopsida C. Aghard	Equisetopsida C. Aghard	<a href="http://dx.doi.org">http://dx.doi.org</a>
3	26		73 Equisetopsida C. Aghard	26	Equisetidae Warming	Equisetidae Warming	<a href="http://dx.doi.org">http://dx.doi.org</a>
4	25		26 Equisetidae Warming	25	Equisetales de Candolle ex Berchtold & J. Presl	Equisetales de Candolle ex Berchtold & J. Presl	<a href="http://www.jstor">http://www.jstor</a>
5	128		25 Equisetales de Candolle ex Berchtold & J. Presl	128	Equisetaceae Michaux ex de Candolle	Equisetaceae Michaux ex de Candolle	<a href="http://www.jstor">http://www.jstor</a>
6	1142		128 Equisetaceae Michaux ex de Candolle	1142	Equisetum Linnaeus	Equisetum Linnaeus	<a href="http://www.eflor">http://www.eflor</a>
7	2004		1142 Equisetum Linnaeus	2004	Equisetum subg. Equisetum	Equisetum subg. Equisetum	<a href="http://www.eflor">http://www.eflor</a>
8	5467		2004 Equisetum subg. Equisetum	5467	Equisetum fluviatile Linnaeus	Equisetum fluviatile	
9	5466		2004 Equisetum subg. Equisetum	5466	Equisetum arvense Linnaeus	Equisetum arvense	
10	5472		2004 Equisetum subg. Equisetum	5472	Equisetum pratense Ehrhart	Equisetum pratense	
11	5471		2004 Equisetum subg. Equisetum	5471	Equisetum palustre Linnaeus	Equisetum palustre	
12	5474		2004 Equisetum subg. Equisetum	5474	Equisetum sylvaticum Linnaeus	Equisetum sylvaticum	
13	5482		2004 Equisetum subg. Equisetum	5482	Equisetum x-litorale K�hlew�in ex Ruprecht	Equisetum x-litorale	
14	5476		2004 Equisetum subg. Equisetum	5476	Equisetum telmateia Ehrhart	Equisetum telmateia	
15	15836		5476 Equisetum telmateia Ehrhart	15836	Equisetum telmateia subsp. braunii (J. Milde) H	Equisetum telmateia	

*“Каталог или лист имен организмов, или таксонов.”*

Также возможны: местные названия, библиография, ...

Как правило, информация классифицируется по таксономическому, географическому или тематическому признаку по этим критериям по отдельности или в некоторой их комбинации.



# Типы данных: находки (occurrences)

	A	B	C	D	E	F	G	H	I	J	K
1	occurrenceID	basisOfRecord	eventDate	endDayOfYear	year	month	day	verbatimEventDate	eventRemarks	scientificName	higherClassific
2	http://arctos.database.m	PreservedSpecimen	1926-04		1926	4		0/4/1926	day of month unknown	Ambystoma maculatum	Animalia; Chorc
3	http://arctos.database.m	PreservedSpecimen	1942-04-17	107	1942	4	17	17/04/1942		Desmognathus fuscus	Animalia; Chorc
4	http://arctos.database.m	PreservedSpecimen	1942-04-17	107	1942	4	17	17/04/1942		Gyrinophilus porphyriticus	Animalia; Chorc
5	http://arctos.database.m	PreservedSpecimen	1942-04-17	107	1942	4	17	17/04/1942		Eurycea bislineata bislineata	Animalia; Chorc
6	http://arctos.database.m	PreservedSpecimen	1942-04-17	107	1942	4	17	17/04/1942		Plethodon cinereus	Animalia; Chorc
7	http://arctos.database.m	PreservedSpecimen	1953-09-27	270	1953	9	27	27-sept-53		Rana sylvatica	Animalia; Chorc
8	http://arctos.database.m	PreservedSpecimen	1979-06-02/1979-06-07					02/06/1979		Eleutherodactylus eneidae	Animalia; Chorc
9	http://arctos.da		1981-06-01	152	1981	6	1	01-juin-81		Masticophis flagellum piceus	Animalia; Chorc
10	http://arctos.da		2011-06-23	174	2011	6	23	23-juin-11		Rana (Lithobates) clamitans	Animalia; Chorc



Цифровая, текстовая или мультимедийная запись описывающая находку организма, то есть **что, где, когда, как и кем** была зафиксирована информация о находке.

# Тип данных: описания, учёты и сборы (sample-event data)



Фактически это структурированный набор данных о находках, в который добавлены общие характеристики, объединяющие отдельные находки:

- характеристики пробных площадей или маршрутов
- описание методики сборов
- объём проделанной работы
- ....

# Тип данных: описания, учёты и сборы (sample-event data)

Каждый отдельный сбор или описание в терминологии Darwin Core называется “**событием**” (**EVENT**).

Эти данные часто представляют на нескольких листах: данные о самом событии (площадка, трансекта, ловушка, квадрат ...), данные о находках видов, учтенных в каждом описании, данные о переменных и т. д.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	eventID	samplingProtocol	samplingEffort	sampleSizeValue	sampleSizeUnit	eventDate	eventTime	startDayOfYear	eventRemarks	country	countryCode	locality	locationID
2	994-tr009-s00	Pollard walks	Average of 30 Minutes walk along transect	250 square metre		2012-10-11	09:28:02Z/10:16:02Z	284	No occurrences	Israel	IL	Sde boker reches halukim	tr009-s00
3	3502-tr056-s6	Pollard walks	Average of 30 Minutes walk along transect	250 square metre		2015-10-19	12:25:02Z/13:10:02Z	291		Israel	IL	Nahal Kovshim Beer Sheva	tr056-s6
4	3502-tr056-s9	Pollard walks	Average of 30 Minutes walk along transect	250 square metre		2015-10-19	12:25:02Z/13:10:02Z	291		Israel	IL	Nahal Kovshim Beer Sheva	tr056-s9

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1	eventID	occurrenceID	basisOfRecord	individualCount	organismQuantity	organismQuantityType	occurrenceStatus	scientificName	kingdom	phylum	class	order	family	infraspecificEpithet	taxonRank	recordedBy
2	1382-tr009-s00	1382-tr009-s00-0	HumanObservation	0	0 individuals		absent	Lepidoptera	Animalia	Arthropoda	Insecta	Lepidoptera			order	Eviatar Fingo
3	3502-tr056-s6	3502-tr056-s6-21114	HumanObservation	3	3 individuals		present	Azanus Jesus	Animalia	Arthropoda	Insecta	Lepidoptera	Lycaenidae		species	Zvika Avni
4	3502-tr056-s6	3502-tr056-s6-21126	HumanObservation	1	1 individuals		present	Melitaea trivia	Animalia	Arthropoda	Insecta	Lepidoptera	Nymphalidae		species	Zvika Avni
5	3502-tr056-s6	3502-tr056-s6-21127	HumanObservation	3	3 individuals		present	Deudorix livia	Animalia	Arthropoda	Insecta	Lepidoptera	Lycaenidae		species	Zvika Avni
6	3502-tr056-s6	3502-tr056-s6-21129	HumanObservation	1	1 individuals		present	Azanus ubaldus	Animalia	Arthropoda	Insecta	Lepidoptera	Lycaenidae		species	Zvika Avni
7	3502-tr056-s6	3502-tr056-s6-21132	HumanObservation	1	1 individuals		present	Lycaena thersamon	Animalia	Arthropoda	Insecta	Lepidoptera	Lycaenidae		species	Zvika Avni
8	3502-tr056-s9	3502-tr056-s9-21116	HumanObservation	1	1 individuals		present	Azanus Jesus	Animalia	Arthropoda	Insecta	Lepidoptera	Lycaenidae		species	Zvika Avni
9	3502-tr056-s9	3502-tr056-s9-21122	HumanObservation	1	1 individuals		present	Tarucus balkanica	Animalia	Arthropoda	Insecta	Lepidoptera	Lycaenidae		species	Zvika Avni
10	3502-tr056-s9	3502-tr056-s9-21131	HumanObservation	1	1 individuals		present	Azanus ubaldus	Animalia	Arthropoda	Insecta	Lepidoptera	Lycaenidae		species	Zvika Avni

## Лицензии

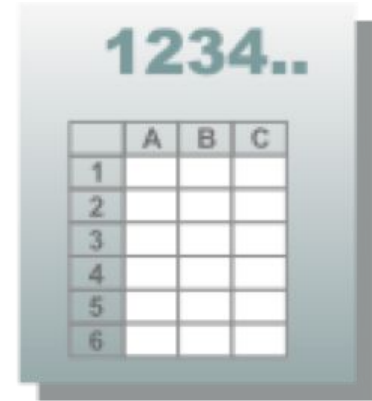
- Сейчас (с 2016) каждый набор данных в GBIF имеет определенную лицензию
- Выбор:
  - Public domain: CC0
  - Creative Commons Attribution: CC-BY
  - Creative Commons Attribution Non Commercial: CC-BY-NC



# Simple Darwin Core и Darwin Core Archive

**простой Darwin Core:** Darwin Core представленный в простой табличной структуре.

включает 169 полей  
поля, для которых значения  
отсутствуют, в таблицу не заносятся



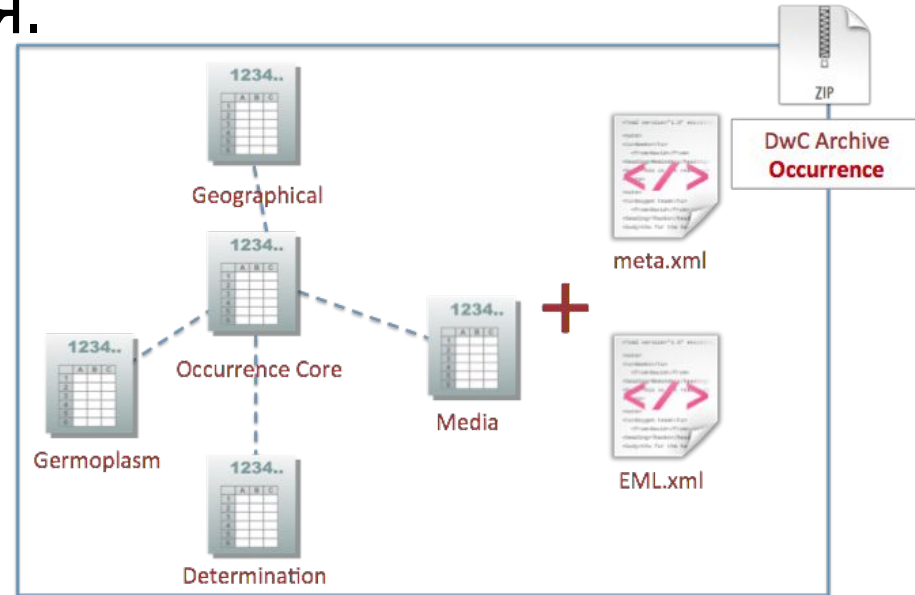


# Darwin Core, простой Darwin Core и Darwin Core архив

**Darwin Core архив:** более комплексный формат, поддерживает расширения.

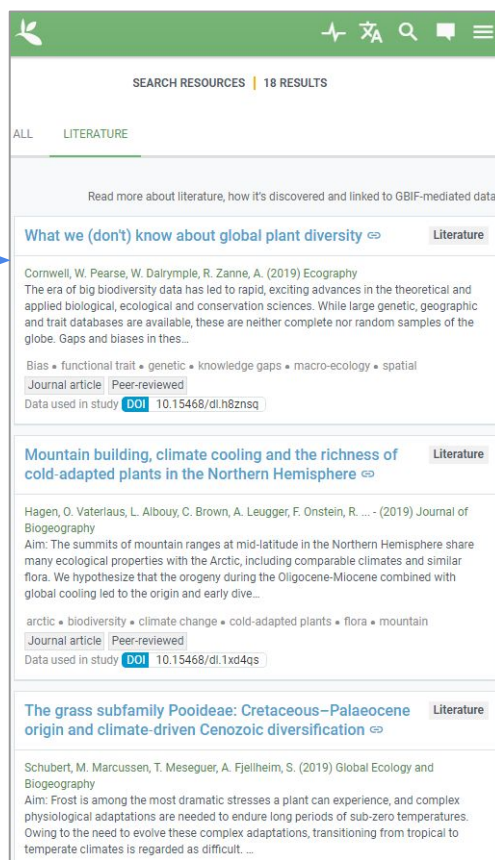
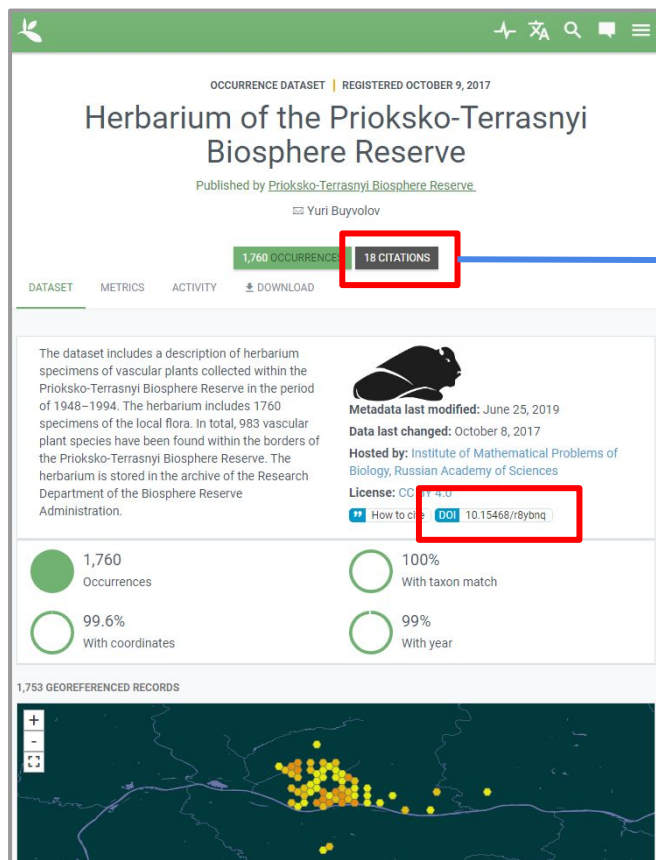
Более 200 возможных полей, в т.ч.

- ссылки на медиа-файлы
- местные названия
- ...





# Идентификатор цифрового объекта (DOI)



DOI присваивается  
автоматически новому  
набору данных в  
процессе регистрации  
в сети GBIF

позволяет  
отслеживать  
использование данных



**СПАСИБО ЗА ВНИМАНИЕ !!!**