

Оцифровка этикеточных данных гербария и их  
подготовка для публикации в GBIF

Практический опыт на примере коллекции  
мохообразных гербария ИБ Коми НЦ УрО РАН  
(SYKO)

*Чадин Иван Федорович (докладчик)*

*Железнова Галина Виссарионовна,  
Литвиненко Галина Александровна,  
Рубцов Михаил Дмитриевич,  
Шубина Татьяна Павловна*

ИБ Коми НЦ УрО РАН

Работа выполнена в рамках проекта, поддержанного  
The Global Biodiversity Information Facility Secretariat  
(GBIFS):

*Data mobilization in European Russia programme contract  
Russia2019\_04:  
European North-East Russia mosses occurrence data mobilization  
on the base of SYKO herbarium mosses collection*

### **Участники проекта**

Чадин Иван Федорович (к.б.н., руководитель проекта, программирование)  
Железнова Галина Виссарионовна (д.б.н., ввод данных, проверка данных)  
Литвиненко Галина Александровна (ввод данных)  
Рубцов Михаил Дмитриевич (определение геогр. координат, ввод данных)  
Шубина Татьяна Павловна (к.б.н., ввод данных, проверка данных)

# Гербарная этикетка как объект для цифровизации



*«Неэтикетированный гербарий – это всего лишь куча сена.»*

*И пока каждый собранный гербарный лист не получил окончательной («чистовой») этикетки, ... – до тех пор опасность стать клочком сена для этого образца не миновала».*

**А. К. Скворцов Гербарий. Пособие по методике и технике. М., Наука, 1977. 199 с.**

# Гербарная этикетка как объект для цифровизации



- 2 экземпляра бумажных этикеток:
- в картотеке этикеток
  - на пакетах (листах) с образцами

# Гербарная этикетка как объект для цифровизации

ГЕРБАРИЙ КОМИ НАУЧНОГО ЦЕНТРА  
УРАЛЬСКОГО ОТДЕЛЕНИЯ РАН (СУКО)

*Dicranum drumondii* Müll. Hal.

Республика Коми: *Княжпогостский р-н,*  
*р. Биваль, приток р. Килговка,*  
*6 км до п. Касяво 27 водок.*

(*62°20' - 50°50'*) кв. 39

*21.07.2010* Leg.: *Тетеряк Б.Ю.*

Det.: *Железнова Г.В.*

№ 12  
№ общий 52388

КОМИ ФИЛИАЛ АН СССР

ГЕРБАРИЙ СПОРОВЫХ РАСТЕНИЙ

*Leskeella radiosa* (Mitt.) Mönket.  
*Fraxinellum caespitum* (Vahl) B.S.P.

Коми АССР *Троицко-Печорский* р-н. Квадр. № 49  
*7 км к Ю-З от р. Стартон, левый берег,*  
*верховья р. Билин.*

Местообитание *памятник* Индекс местооб. *224*  
*русло ручья, у болот.*

*61°51' с.ш. 59°16' в.д.*

№ полевой 163  
№ общий 1303

Собрал  
*10. VII. 1966г.* Г.В. Железнова

Определил

Рис.  
Преп.  
Разм.  
Кач.

КОМИ ФИЛИАЛ АН СССР

ГЕРБАРИЙ СПОРОВЫХ РАСТЕНИЙ

*Fraxinellum caespitum* (Vahl) B.S.P.

Коми АССР, *Воркутинский* район,  
*Г. Воркута, цементный завод* квадрат № 2

Местообитания *Землистый с камнями склон*  
*к р. Воркута*

*12-VII-69* Шифр местооб. 01  
Собрал *Железнова Г.В.* № образца полевой *32/3*

Определил *Железнова Г.В.* общий 5848

Этикетки, изготовленные в разное время немного отличаются по размеру

# Гербарная этикетка как объект для цифровизации

Как ускорить подготовку  
этикеточных данных к публикации?

# Гербарная этикетка как объект для цифровизации

Как ускорить подготовку  
этикеточных данных к публикации?

1. Получение цифровых изображений  
этикеток

# Получение цифровых изображений этикеток

## планшетный сканер



### Преимущества

- высокое качество изображения
- небольшая цена (5–30 тыс. руб.)

### Недостатки

- низкая скорость
- непригодны для хрупких объектов
- непригодны для объемных объектов

*Сканер CANON CanoScan 9000F MARK II  
изображение с сайта citilink.ru/*

# Получение цифровых изображений этикеток

## Готовые сканеры на основе фотоаппаратов



### Преимущества

- высокое качество изображения
- большая производительность
- сканирование хрупких объектов
- сканирование объемных объектов

### Недостатки

- большая стоимость (500 – 1500 тыс. руб.)

Модель ЭларСкан А2 <https://elarscan.ru/modelnyj-ryad/>

# Получение цифровых изображений этикеток

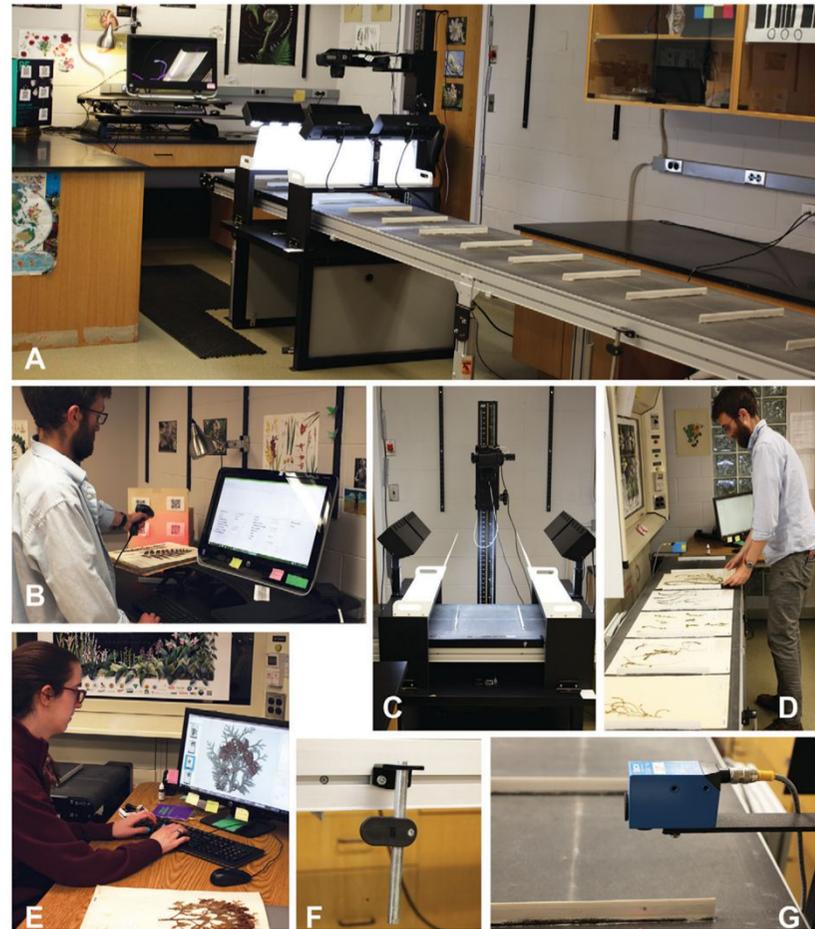
## Самодельные сканеры на основе фотоаппаратов



*Nelson, G. et al. Five task clusters that enable efficient and effective digitization of biological collections / G. Nelson, D. Paul, G. Riccardi, A. Mast // ZooKeys. – 2012. – V. 209. – P. 19–45. – DOI: 10.3897/zookeys.209.3135*

# Получение цифровых изображений этикеток

## Самодельные сканеры на основе фотоаппаратов



**Fig. 7.** Images of conveyor system and sub-components. **A**, View of conveyor system; **B**, Data Entry User Interface (DEUI); **C**, Lighting system and camera; **D**, End of conveyor showing manual removal of specimens; **E**, Main (System Controller) computer quality control interface; **F**, Photo-eye to control advance of conveyor belt; **G**, Contrast sensor to detect specimens at end of belt, preventing specimens from falling off of belt.

Sweeney, P. W. et al. Large-scale digitization of herbarium specimens: Development and usage of an automated, high-throughput conveyor system / P. W. Sweeney, B. Starly, P. J. Morris, Y. Xu, A. Jones, S. Radhakrishnan, C. J. Grassa, C. C. Davis // TAXON. – 2018. – V. 67. – N 1. – P. 165–178. – DOI: 10.12705/671.10

# Получение цифровых изображений этикеток

## Самодельные сканеры на основе фотоаппаратов



Фото с сайта:

<http://digitalphotorepro.blogspot.com/2011/04/standardized-copy-stand-photography-for.html>

### Преимущества

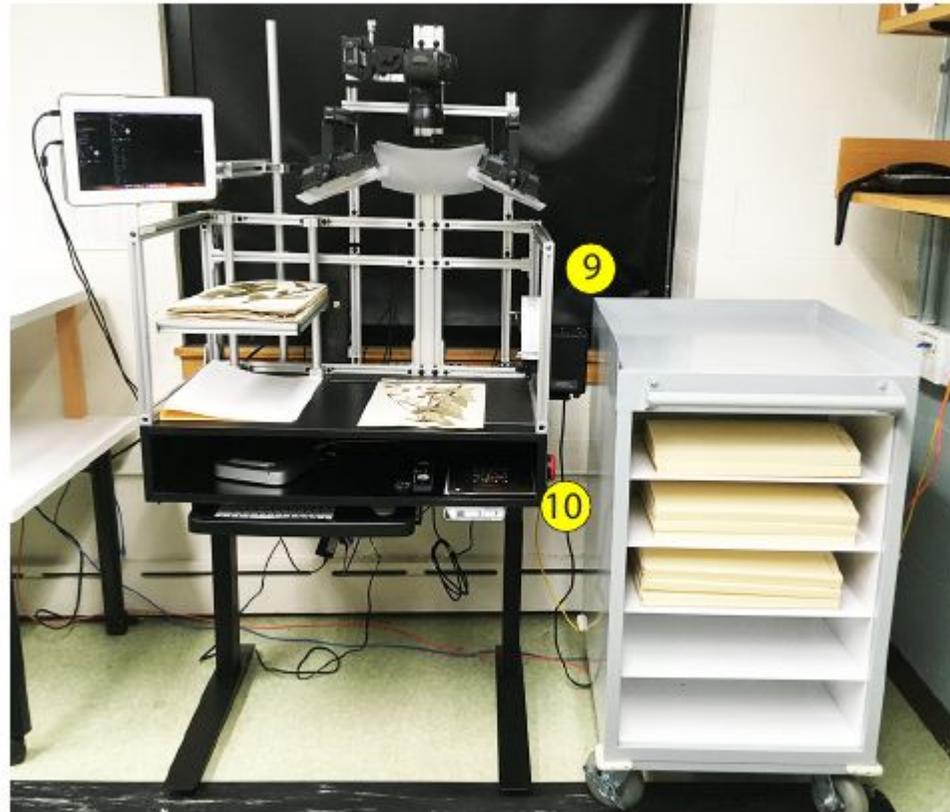
- высокое качество изображения
- большая производительность
- бережное сканирование
- стоимость зависит от марки фотоаппарата (10–350 тыс. руб.)

### Недостатки

- требует ручной сборки и настройки

# Получение цифровых изображений этикеток

## Самодельные сканеры на основе фотоаппаратов



*Back to the future: A refined single-user photostation for massively scaling herbarium digitization*  
Charles C. Davis, Jonathan A. Kennedy, Christopher J. Grassa, First published: 16 March 2021  
<https://doi.org/10.1002/tax.12459>

# Получение цифровых изображений этикеток

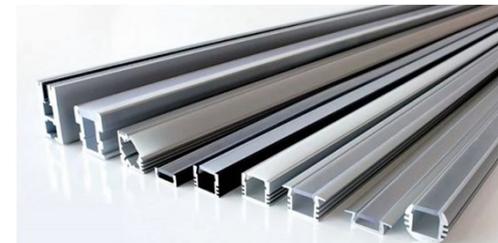
## Наш самодельный сканер на основе фотоаппарата



Фотоаппарат Canon PowerShot A480



Светодиодные лампы

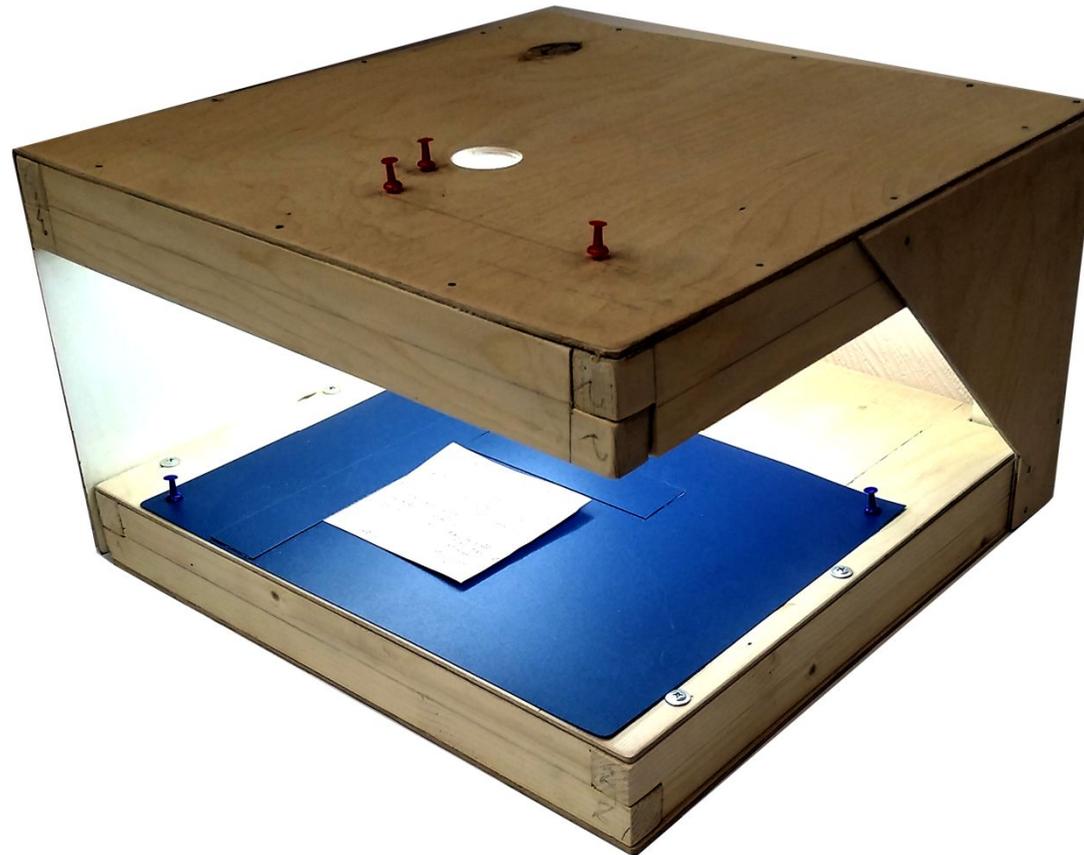


Алюминиевый профиль для мебели

**5-7 секунд на 1 этикетку**

# Получение цифровых изображений этикеток

**Наш самодельный сканер на основе смартфона**



**3-4 секунды на 1 этикетку**

# Фотография этикетки до обработки

ГЕРБАРИЙ КОМИ НАУЧНОГО ЦЕНТРА  
УРАЛЬСКОГО ОТДЕЛЕНИЯ РАН (СУКО)

*Isopterygiopsis pulchella* (Niedw.) Iwats.

Коми SSR Княжпогостский р-н,  
лев. бер. р. Вымь, окр. ретрансляцион-  
ной вышки. Смешан. кустарничково-  
хвощевой мелкотравно-моховой.

На выворотине. (81) об

Квадр. № 39

10 06<sup>19</sup> 94

Железнова Г.В.  
Leg. Шубина Т.П.

№ 25

Determ. Шубина Т.П.

№ общий 27767

Фотография этикетки после автоматического отсечения фона

ГЕРБАРИЙ КОМИ НАУЧНОГО ЦЕНТРА  
УРАЛЬСКОГО ОТДЕЛЕНИЯ РАН (СУКО)

*Isoterygiopsis pulchella (Nied.) T. Wats.*

Коми SSR Княжпогостский р-н,  
лев. бер. р. Вымь, окр. ретрансляцион-  
ной вышки. Смешан. кустарничково-  
хвощевой мелкотравно-моховой.

На выворотине. (81) об

Квадр. № 39

Железнова Г. В.  
Leg. Шубина Т. П.  
Determ. Шубина Т. П.

10 06<sup>19</sup> 94  
№ 25  
№ общий 27767

# Гербарная этикетка как объект для цифровизации

Какой минимум данных *желателен*  
для регистрации находки в GBIF?

# Гербарная этикетка как объект для цифровизации

Какой минимум данных желателен для регистрации находки в GBIF

ГЕРБАРИЙ КОМИ НАУЧНОГО ЦЕНТРА  
УРАЛЬСКОГО ОТДЕЛЕНИЯ РАН (СУКО)

*Dicranum drumondii Müll. Hal.*

Республика Коми: Княжпогожевский р-н,  
р. Бичваль, приток р. Килёвка,  
6 км до п. Касёво 3 водок.

(62°20' - 50°50')

21.07.2010

№ 12

№ общий 52388

Лсг.: Тетерюк Б.Ю.

Det.: Железнова Т.В.

Кв. 39

# Классификация данных этикетки по признакам сложности ввода и их важности для описания находки вида



# Гербарная этикетка как объект для цифровизации

Как ускорить подготовку  
этикеточных данных к публикации?

# Типичные хранилища данных



и другие реляционные СУБД

*Когда коллег больше чем 2*

*Когда записей больше 10 тыс.*



и другие эл. таблицы

*Когда вас не больше двух*

*Вам хватает*

*1 048 576 строк*

*и 16 384 столбца*

# Ввод данных в электронную таблицу

ГЕРБАРИЙ КОМИ НАУЧНОГО ЦЕНТРА  
УРАЛЬСКОГО ОТДЕЛЕНИЯ РАН (СУКО)

*Dicranum drumondii Müll. Hal.*

Республика Коми: Княжпогощенский р-н,  
р. Бываль, приток р. Килговка,  
6 км до п. Килгово З. берег.

(62°20' - 50°50') кв. 39

29.07.2010 Лсг.: Тетерук Б.Ю.

№ 12 Дст.: Железнова Г.В.  
№ общий 52388

Для\_примера\_May\_2019\_For\_Pub\_GBIF.xls - LibreOffice Calc

Файл Правка Вид Вставка Формат Лист Данные Сервис Окно Справка

Arial 10

B4993 Dicranum

	A	B	C	D	E	F	G	H
1	catalogNum	scientificName	recordedBy	identifiedBy	associatedMedia	day	month	year
4972	55497	Sphagnum balticum (Russ.) C. Jer	Kulyugina E.E.	Zheleznova G.V.	http://ib.komisc.	27	7	201
4973	54936	Sphagnum capillifolium (Ehrh.) Her	Dulin M.V.	Zheleznova G.V.	http://ib.komisc.	8	7	201
4974	55077	Sphagnum capillifolium (Ehrh.) Her	Dulin M.V.	Zheleznova G.V.	http://ib.komisc.	17	7	201
4975	55315	Sphagnum capillifolium (Ehrh.) Her	Kulyugina E.E.	Zheleznova G.V.	http://ib.komisc.	10	7	201
4976	55333	Sphagnum capillifolium (Ehrh.) Her	Kulyugina E.E.	Zheleznova G.V.	http://ib.komisc.	10	7	201
4977	54898	Sphagnum cuspidatum Ehrh. ex H	Dulin M.V.	Zheleznova G.V.	http://ib.komisc.	6	7	201
4978	54950	Sphagnum cuspidatum Ehrh. ex H	Dulin M.V.	Zheleznova G.V.	http://ib.komisc.	10	7	201
4979	55142	Sphagnum cuspidatum Ehrh. ex H	Kulyugina E.E.	Zheleznova G.V.	http://ib.komisc.	28	6	201
4980	55142	Straminergon stramineum Hedenä	Kulyugina E.E.	Zheleznova G.V.	http://ib.komisc.	28	6	201
4981	55661	Sphagnum cuspidatum Ehrh. ex H	Teteryuk B.Yu.	Zheleznova G.V.	http://ib.komisc.	13	7	201
4982	56317	Sphagnum angustifolium (C. Jens)	Teteryuk L.V.	Zheleznova G.V.	http://ib.komisc.	8	8	201
4983	56317	Sphagnum girgensohnii Russ.	Teteryuk L.V.	Zheleznova G.V.	http://ib.komisc.	8	8	201
4984	56329	Sphagnum capillifolium (Ehrh.) Her	Teteryuk L.V.	Zheleznova G.V.	http://ib.komisc.	19	6	201
4985	57122	Sphagnum angustifolium (C. Jens)	Kulyugina E.E.	Zheleznova G.V.	http://ib.komisc.	11	7	201
4986	57073	Sphagnum capillifolium (Ehrh.) Her	Kulyugina E.E.	Zheleznova G.V.	http://ib.komisc.	7	7	201
4987	57073	Tomentypnum nitens (Hedw.) Loes	Kulyugina E.E.	Zheleznova G.V.	http://ib.komisc.	7	7	201
4988	57085	Sphagnum capillifolium (Ehrh.) Her	Kulyugina E.E.	Zheleznova G.V.	http://ib.komisc.	8	7	201
4989	57176	Paludella squarrosa (Hedw.) Brid.	Kulyugina E.E.	Zheleznova G.V.	http://ib.komisc.	16	7	201
4990	57176	Sphagnum capillifolium (Ehrh.) Her	Kulyugina E.E.	Zheleznova G.V.	http://ib.komisc.	16	7	201
4991	57176	Sphagnum warnstorffii Russ.	Kulyugina E.E.	Zheleznova G.V.	http://ib.komisc.	16	7	201
4992	57181	Sphagnum capillifolium (Ehrh.) Her	Kulyugina E.E.	Zheleznova G.V.	http://ib.komisc.	16	7	201
4993	52388	Dicranum						
4994								
4995								
4996								



**Преимущества:** – быстрое начало работы

**LibreOffice**



# Ввод данных в электронную таблицу

ГЕРБАРИЙ КОМИ НАУЧНОГО ЦЕНТРА  
УРАЛЬСКОГО ОТДЕЛЕНИЯ РАН (СУКО)

*Dicranum drumondii Müll. Hal.*

Республика Коми: Княжпогожевский р-н,  
р. Бываль, приток р. Килговка,  
6 км до п. Килгово З. берег.

(62°20' - 50°50') кв. 39

29.07.2010 Leg.: Тетерук Б.Ю.

№ 12 Det.: Железнова Г.В.  
№ общий 52388

Для\_примера\_May\_2019\_For\_Pub\_GBIF.xls - LibreOffice Calc

Файл Правка Вид Вставка Формат Лист Данные Сервис Окно Справка

Arial 10

B4993 Dicranum

	A	B	C	D	E	F	G	H
1	catalogNum	scientificName	recordedBy	identifiedBy	associatedMedia	day	month	year
4972	55497	Sphagnum balticum (Russ.) C. Jer	Kulyugina E.E.	Zheleznova G.V.	http://ib.komisc.	27	7	201
4973	54936	Sphagnum capillifolium (Ehrh.) Her	Dulin M.V.	Zheleznova G.V.	http://ib.komisc.	8	7	201
4974	55077	Sphagnum capillifolium (Ehrh.) Her	Dulin M.V.	Zheleznova G.V.	http://ib.komisc.	17	7	201
4975	55315	Sphagnum capillifolium (Ehrh.) Her	Kulyugina E.E.	Zheleznova G.V.	http://ib.komisc.	10	7	201
4976	55333	Sphagnum capillifolium (Ehrh.) Her	Kulyugina E.E.	Zheleznova G.V.	http://ib.komisc.	10	7	201
4977	54898	Sphagnum cuspidatum Ehrh. ex H	Dulin M.V.	Zheleznova G.V.	http://ib.komisc.	6	7	201
4978	54950	Sphagnum cuspidatum Ehrh. ex H	Dulin M.V.	Zheleznova G.V.	http://ib.komisc.	10	7	201
4979	55142	Sphagnum cuspidatum Ehrh. ex H	Kulyugina E.E.	Zheleznova G.V.	http://ib.komisc.	28	6	201
4980	55142	Straminergon stramineum Hedenä	Kulyugina E.E.	Zheleznova G.V.	http://ib.komisc.	28	6	201
4981	55661	Sphagnum cuspidatum Ehrh. ex H	Teteryuk B.Yu.	Zheleznova G.V.	http://ib.komisc.	13	7	201
4982	56317	Sphagnum angustifolium (C. Jens)	Teteryuk L.V.	Zheleznova G.V.	http://ib.komisc.	8	8	201
4983	56317	Sphagnum girgensohnii Russ.	Teteryuk L.V.	Zheleznova G.V.	http://ib.komisc.	8	8	201
4984	56329	Sphagnum capillifolium (Ehrh.) Her	Teteryuk L.V.	Zheleznova G.V.	http://ib.komisc.	19	6	201
4985	57122	Sphagnum angustifolium (C. Jens)	Kulyugina E.E.	Zheleznova G.V.	http://ib.komisc.	11	7	201
4986	57073	Sphagnum capillifolium (Ehrh.) Her	Kulyugina E.E.	Zheleznova G.V.	http://ib.komisc.	7	7	201
4987	57073	Tomentypnum nitens (Hedw.) Loes	Kulyugina E.E.	Zheleznova G.V.	http://ib.komisc.	7	7	201
4988	57085	Sphagnum capillifolium (Ehrh.) Her	Kulyugina E.E.	Zheleznova G.V.	http://ib.komisc.	8	7	201
4989	57176	Paludella squarrosa (Hedw.) Brid.	Kulyugina E.E.	Zheleznova G.V.	http://ib.komisc.	16	7	201
4990	57176	Sphagnum capillifolium (Ehrh.) Her	Kulyugina E.E.	Zheleznova G.V.	http://ib.komisc.	16	7	201
4991	57176	Sphagnum warnstorffii Russ.	Kulyugina E.E.	Zheleznova G.V.	http://ib.komisc.	16	7	201
4992	57181	Sphagnum capillifolium (Ehrh.) Her	Kulyugina E.E.	Zheleznova G.V.	http://ib.komisc.	16	7	201
4993	52388	Dicranum						
4994								
4995								
4996								



## Недостатки:

- не поддается распараллеливанию
- нельзя использовать контроль ввода
- постоянное переключение внимания

# Типичные хранилища данных



и другие реляционные СУБД

*Когда коллег больше чем 2*

*Когда записей больше 10 тыс.*



и другие эл. таблицы

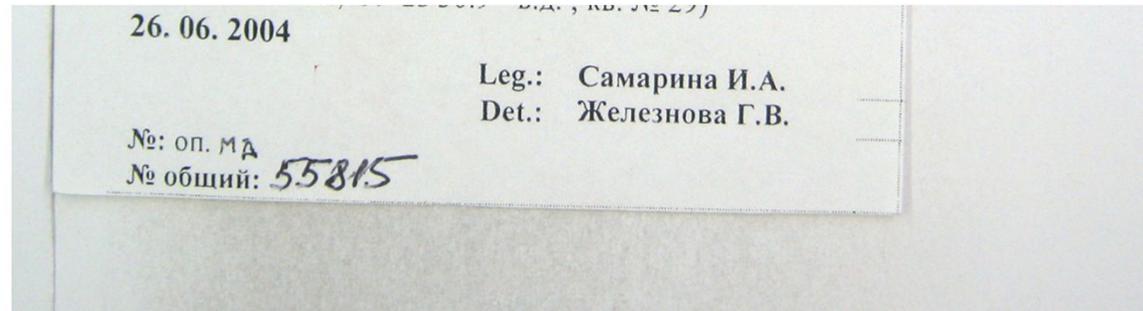
*Когда вас не больше двух*

*Вам хватает*

*1 048 576 строк*

*и 16 384 столбца*

# Ввод данных в СУБД



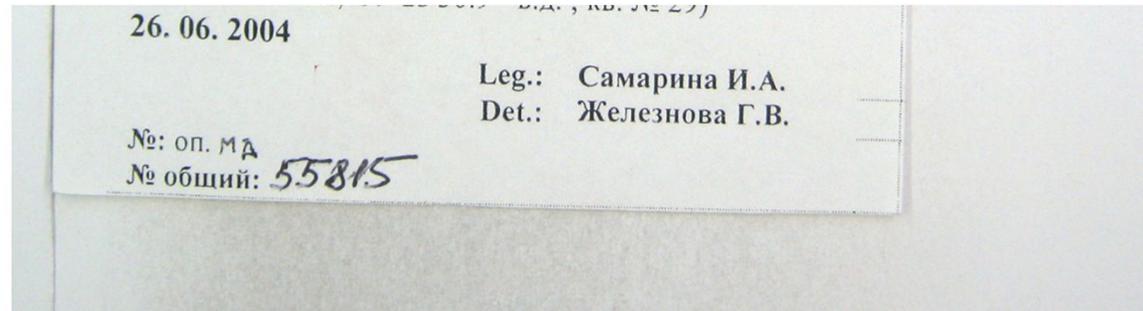
ID:	11924						
Вид:	× Paraleucobryum longifolium +						
День:	26	Месяц:	6	Год:	2004	Catalog num:	55815
Собрал:	× Самарина И. А. +		Определил:	× 1_Железнова Г.В. +			
Ш, Г:	63	Ш, М:	54	Ш, С:	8,6		
Д, Г:	59	Д, М:	23	Д, С:	30,9		
Требует правки:	Нет						
Требует правки даты:	Нет						

## Преимущества:



- поддается распараллеливанию
- гибкие возможности контроля ввода
- можно концентрировать внимания оператора

# Ввод данных в СУБД



ID:	11924						
Вид:	× Paraleucobryum longifolium +						
День:	26	Месяц:	6	Год:	2004	Catalog num:	55815
Собрал:	× Самарина И. А. +		Определил:	× 1_Железнова Г.В. +			
Ш, Г:	63	Ш, М:	54	Ш, С:	8,6		
Д, Г:	59	Д, М:	23	Д, С:	30,9		
Требуется правки:	Нет						
Требуется правки даты:	Нет						

## Недостатки:

- требует разработки структуры базы данных
- требует разработки интерфейса ввода данных



# Демонстрация работы по вводу этикеток

# Рекомендации по организации работы по вводу этикеточных данных

1. Получить цифровые изображения этикеток

# Рекомендации по организации работы по вводу этикеточных данных

2. Разработать веб-интерфейс к базе данных

# Рекомендации по организации работы по вводу этикеточных данных

3. Разделить данные этикеток на три группы:
- **первая очередь**: вид, дата, сборщик, детерминатор
  - **вторая очередь**: координаты места сбора
  - **третья очередь**: текстовое описание места сбора, данные о местообитании, субстрате

# Рекомендации по организации работы по вводу этикеточных данных

4. Для каждой «очереди» разработать серию форм ввода данных. Например, для первой очереди:

- форма для ввода элементов даты;
- форма для ввода видов;
- форма для ввода имен сборщиков и детерминаторов.

5. Дать оператору возможность копировать данные из ранее обработанных этикеток

6. Ввод координат выполнять ПОСЛЕ

а) группировки этикеток по:

- году сбора,

- Ф.И.О. сборщика,

б) сортировки отобранной группы этикеток по дате сбора

7. Каждая форма должна быть снабжена счетчиками:

- сколько сделано сегодня
- сколько сделано всего
- сколько осталось сделать

8. Разработать несколько специализированных форм для проверки введенных данных, разбив проверяемые данные на однородные группы



Фредерик Тейлор  
1856–1915

