

Data Refactoring for Amateurs

Avi Bryant

“intelligent computer users
who [are] not computer
programmers or software
developers in any sense”

–Guido

Welcome to SMALLTALK [May 5]

□ 3 + 4!

7

□ for i←1 to 300 do (@ go i turn 89)!

□ to fact n (if :n=0 then 1 else (n*fact n-(1)))!

fact

□ fact 5!

120

□ edit fact!

edited

□

Behavior

Browser Edit Find View Package Class Protocol Method Tools Help

Find:

Package Hierarchy Instance Class Shared Variable Instance Variable

Local Image
 Base VisualWc
 Kernel
 System *
 Collections
 External *
 Graphics *

AnnotatedMetho
Behavior
 BehaviorBuilderI
 BehaviorShell
 BlockClosure
 BlockContext
 Boolean

initialize-release
 instance creation
 accessing
 accessing class hiera
 accessing instances
 accessing method dic
 binary storage

Source Comment Hierarchy Diagram Rewrite Code Critic Definition

Behavior

super	Core.Object
indexed type	none
category	Kernel-Classes
namespace	Smalltalk.Core
visibility	public
imports	

Accept Cancel

superclass <Behavior> links the chain of inheritance
methodDict <MethodDictionary> associates message names with methods
format <Integer> encodes storage layout of instances
subclasses <Array> back-pointers to the class's subclasses

Class: Core.Behavior **Parcel:** none **Package:** Kernel-Classes

Project-A



"send Boy 2 messages
to make him happy"
Boy face grin; happy

Project-B

Project-B 1 level deeper

Project-C 2 levels below Project-A

Workspace

This poor Workspace is rotating with 6 degrees per second, forced by the scripting tile below it. Fortunately this does not hamper its function

Workspace script1 paused

Workspace turn by 6

Debug

a SystemWindow(159)

- root: a SystemWindow(1592)
- bounds: 314@116 corner: 580@299
- owner: a PasteUpMorph(3926) [world]
- submorphs: #(a ProjectViewMorph(3926))
- fullBounds: 314@116 corner: 580@299
- color: (Color r: 0.979 g: 0.94 b: 0.862)
- extension: a MorphExtension (686) [e: 1]
- borderWidth: 2
- borderColor: #raised
- model: a Project(Project-B)
- slotName: nil

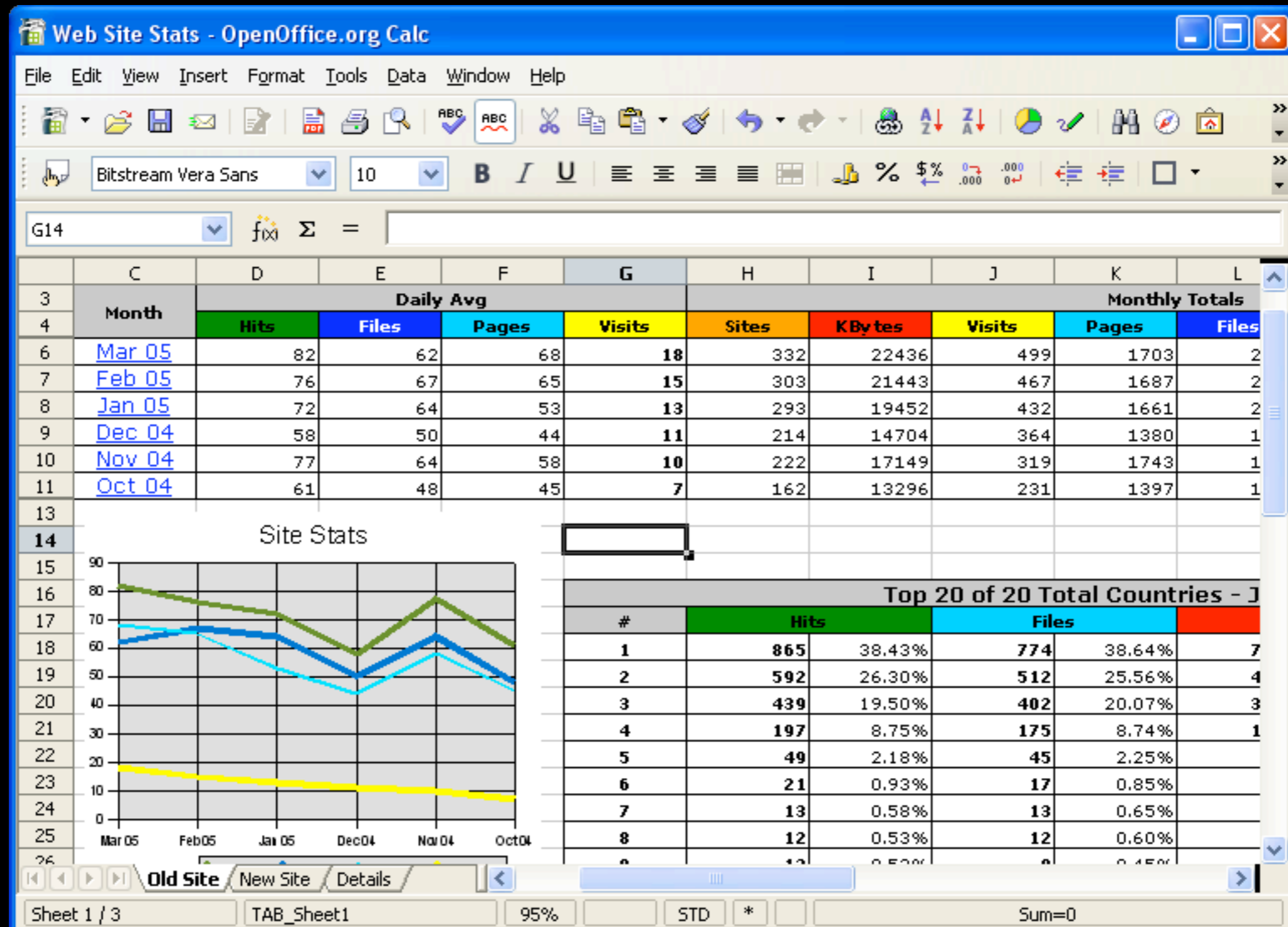
"programming interactively = sending a message to any object and getting an answer"

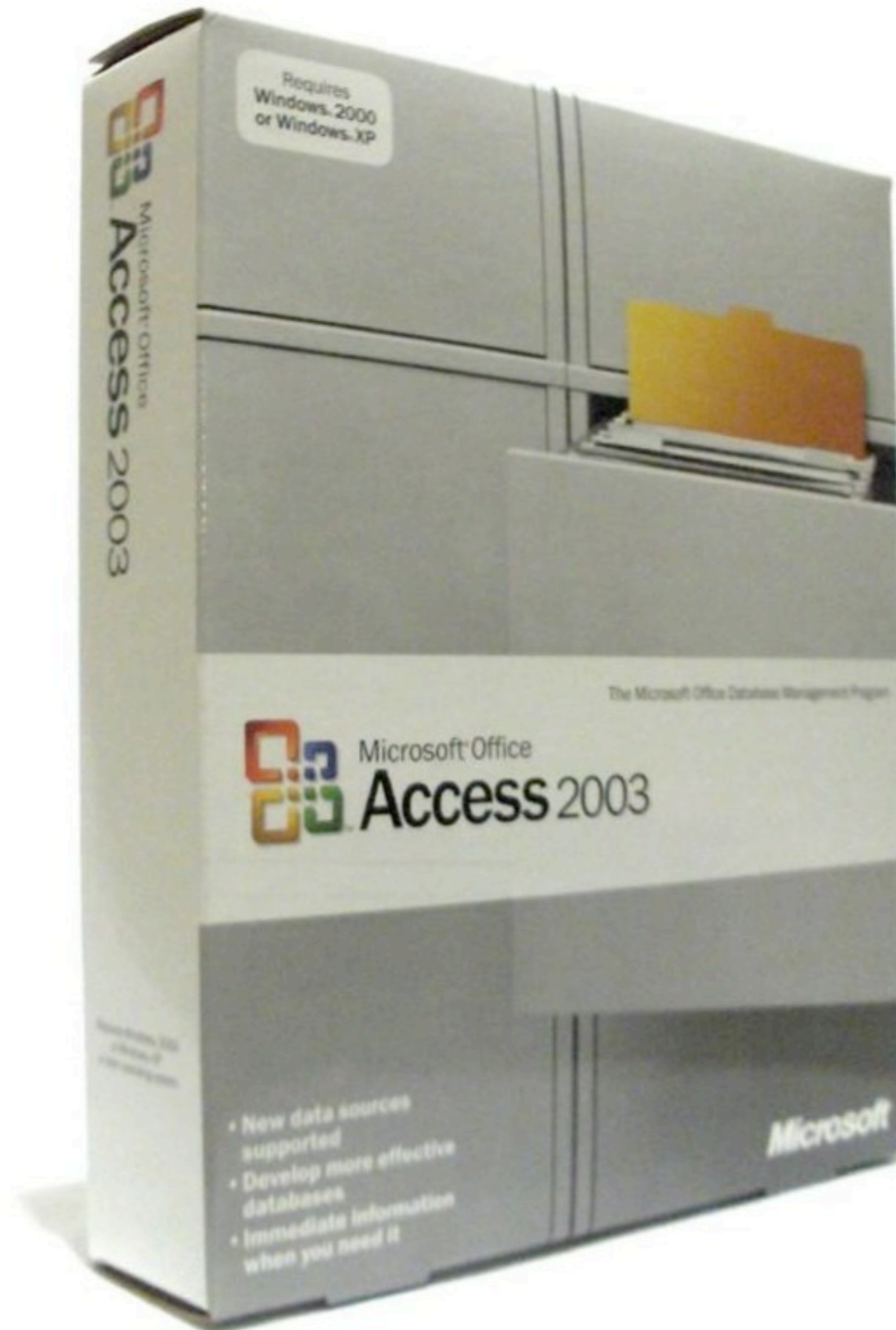
```
self subProjects  
  #(a Project(Project-C))  
self name  
  'Project-B'
```

Interactively created data must be migrated, not recreated.

End-user data is always interactively created.

Interactive and incremental spreadsheets





vs. BDUF databases

The Curse of Legacy Data

How can we make legacy data into a blessing?

Making old data do new things is extremely satisfying.

Painkillers and Vitamins

Users will refactor data to:

- Take away pain when entering new data
- Examine old data in more and better ways

Event	Type	Time	Length	Speakers
Openness and simplicity in dynamic systems implementation	Invited Talk	8:30 am	1hr	Ian Piumarta
PyPy's Approach to Virtual Machine Construction	Research Paper	10am	30 min	Armin Rigo, Samuele Pedroni
Hop, a Language for Programming the Web 2.0	Research Paper	14:30	30min	Manuel Serrano, Erick Gallesio, Florian Loitsch
Dynamic Languages for Statically-Typed Minds	Tutorial	8:30 am	1:30	Niclas Nilsson

Assign Type

You have scalar data in a text field. You cannot sort, filter, or summarize it correctly.

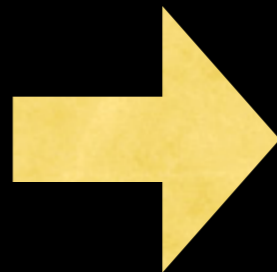
Change the type of the field to match the data.

Invalid: any text that cannot be parsed by the new type.

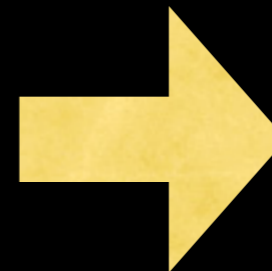
Also: Assign Type with Context

Assign Type

Length (Text)
1hr
30 min
30min
1:30



Length (Duration)
1:00:00
0:30:00
0:30:00
1:30:00

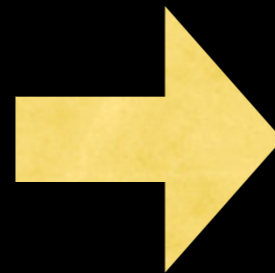


Length (Duration)
0:30:00
0:30:00
1:00:00
1:30:00
3:30:00

Assign Type with Context

Time (Text)
8:30 am
10 am
14:30
8:30 am

Default:
Oct 23, 2006

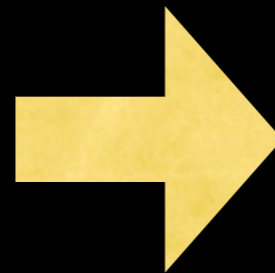


Time (Date/Time)
Oct 23, 2006 8:30 am
Oct 23, 2006 10:00 am
Oct 23, 2006, 2:30 pm
Oct 23, 2006 8:30 am

Assign Type with Context

Cost (Text)
123.00
57.00 CAD
43.50

Default:
USD



Cost (Money)
123.00 USD
57.00 CAD
43.50 USD

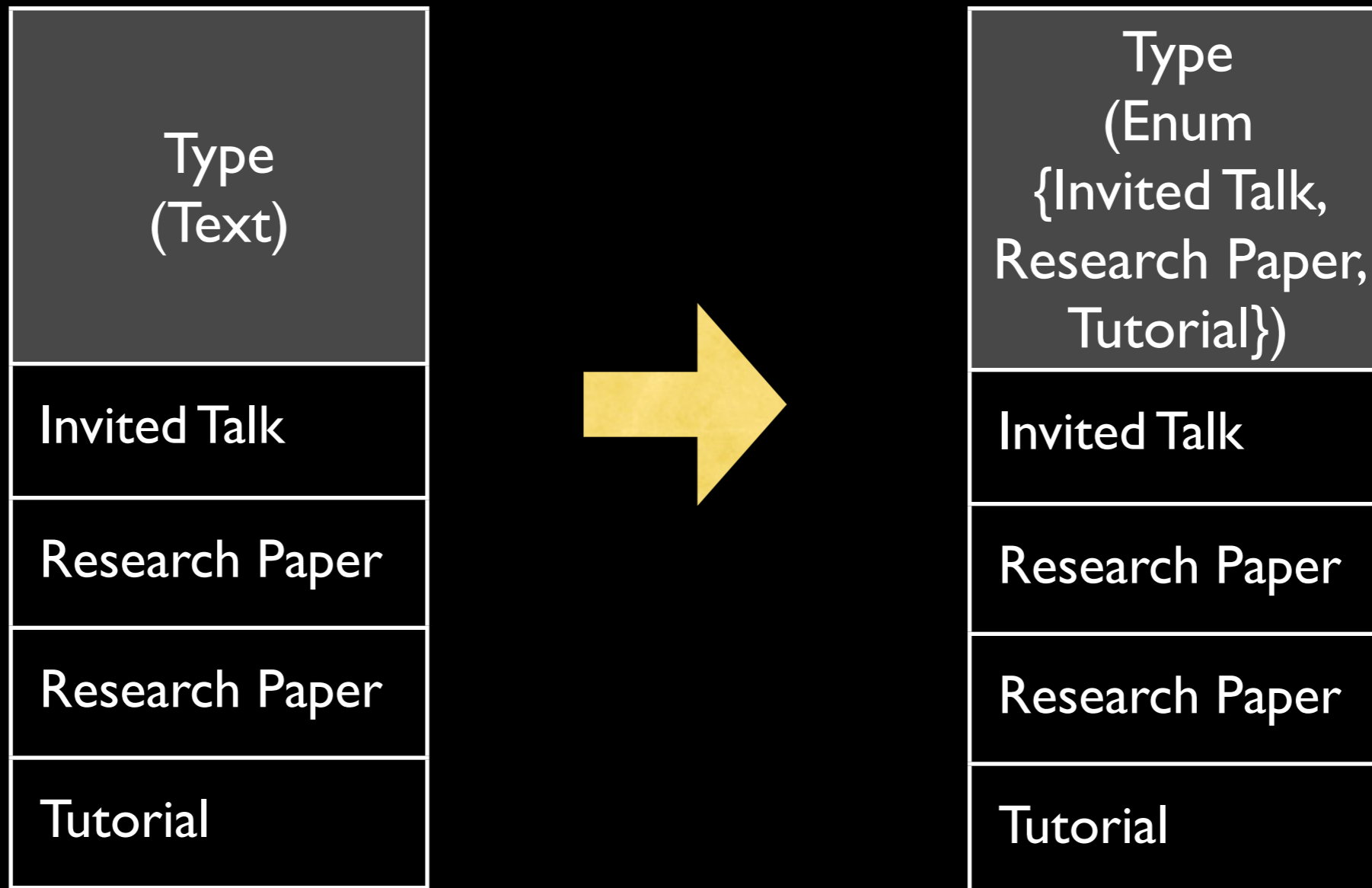
Limit Options

You have a text field with a small number of values that are each used many times. Data entry is inefficient and error prone.

Migrate it to an enumerated type which can be represented as a menu, radio buttons, etc.

Each of the currently used values becomes an option in the new field.

Limit Options



A **great** data model is a good specification
for an **adequate** user interface.

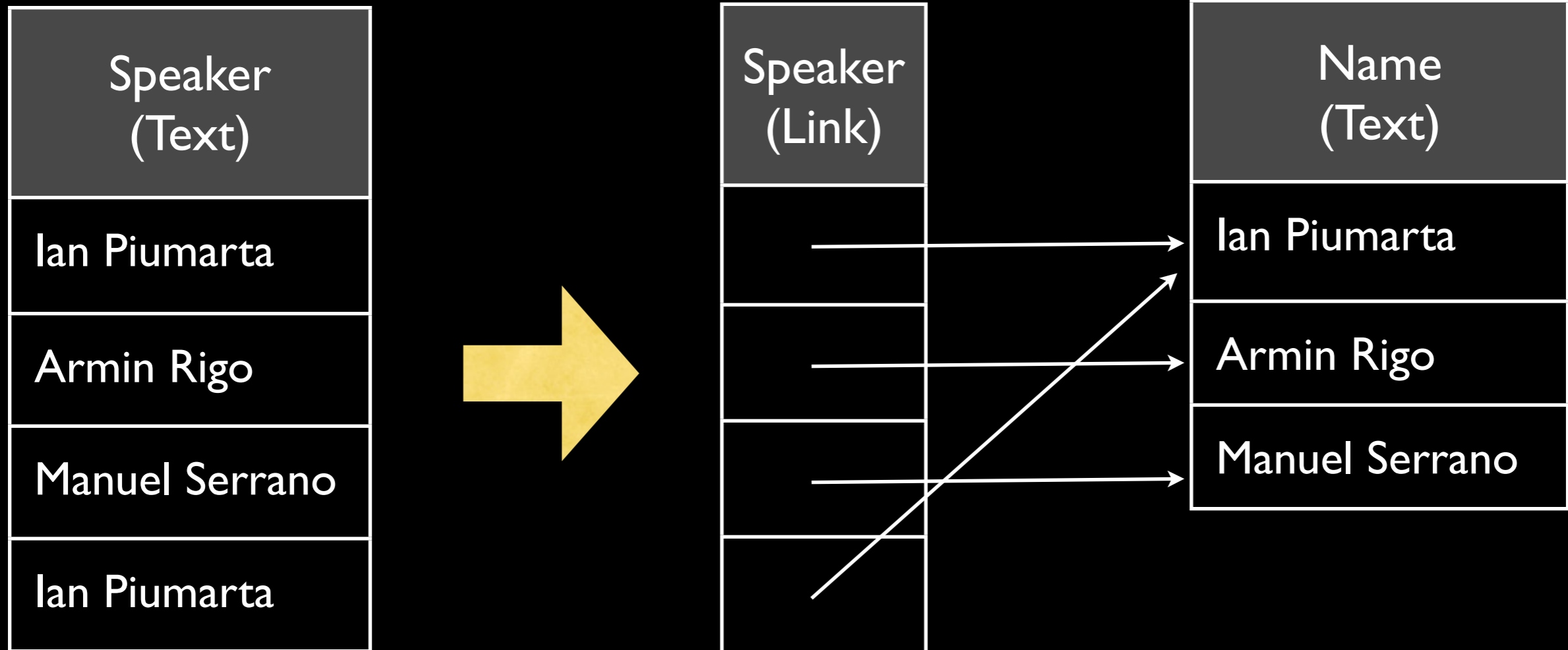
Extract Link

You have a text field where the same value may appear more than once. You want to attach extra information to each of those values.

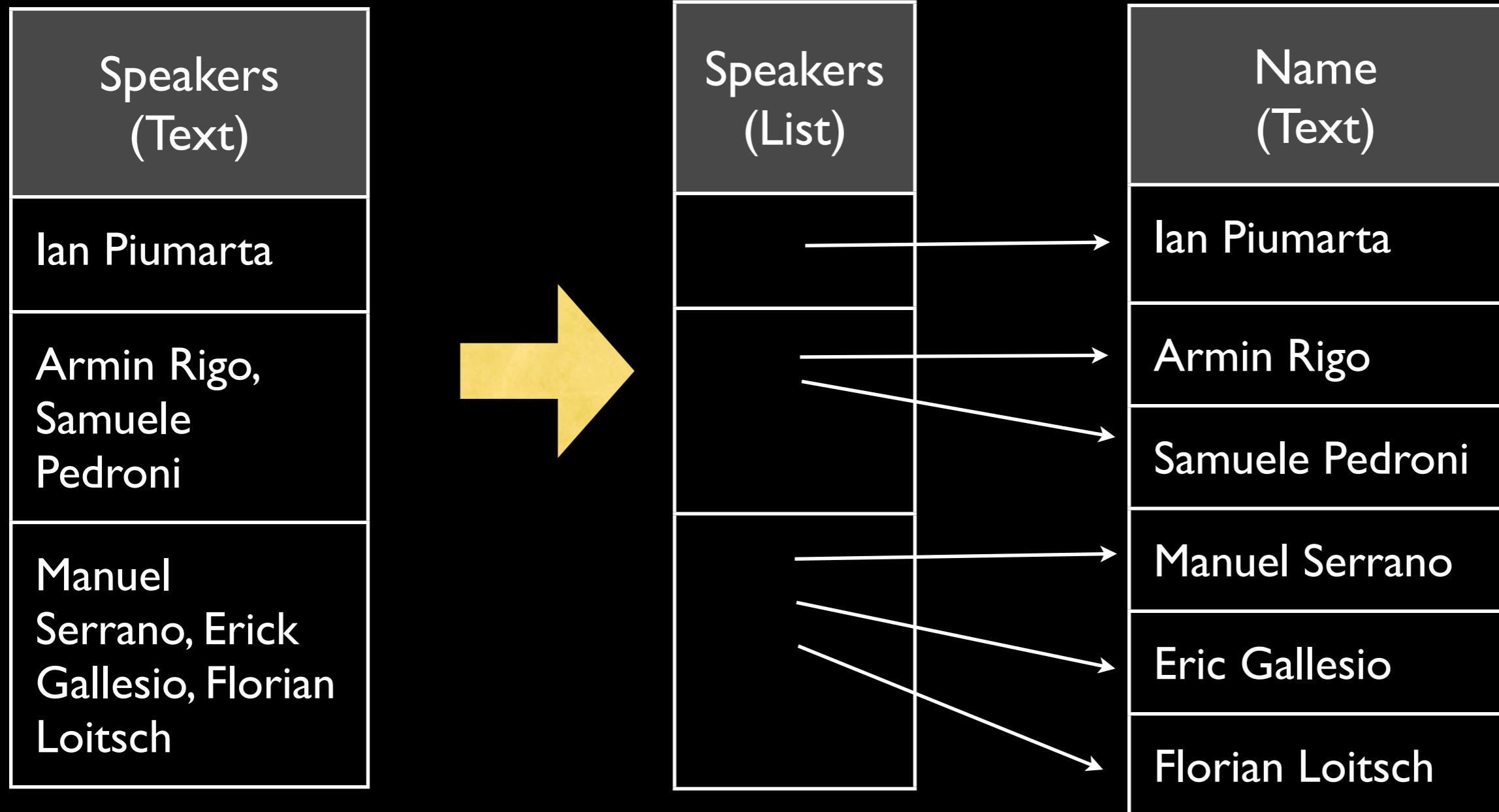
Create a new table containing those values, and replace the text field with links to that new table's rows. Only one new row should be created for each unique value.

Also: Extract List

Extract Link



Extract List



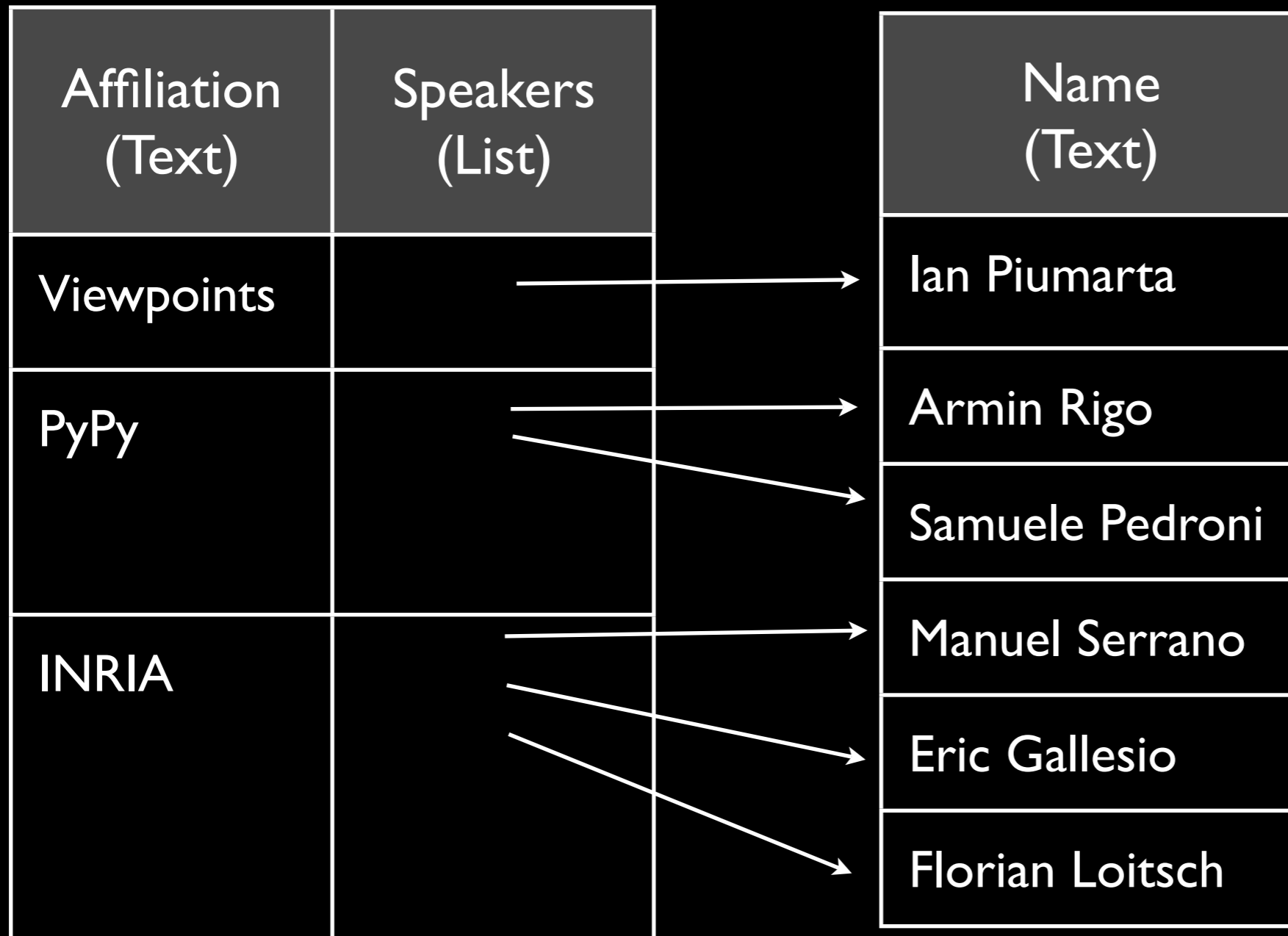
Event	Time	Length	Speaker 1	Affiliation	Speaker 2	Affiliation
Openness and simplicity in dynamic systems implementation	8:30 am	1hr	Ian Piumarta	Viewpoints		
PyPy's Approach to Virtual Machine Construction	10am	30 min	Armin Rigo	DHHU	Samuele Pedroni	ETH
Dynamic Languages for Statically-Typed Minds	8:30 am	1:30	Niclas Nilsson	Activa		

Move Field to Related Table

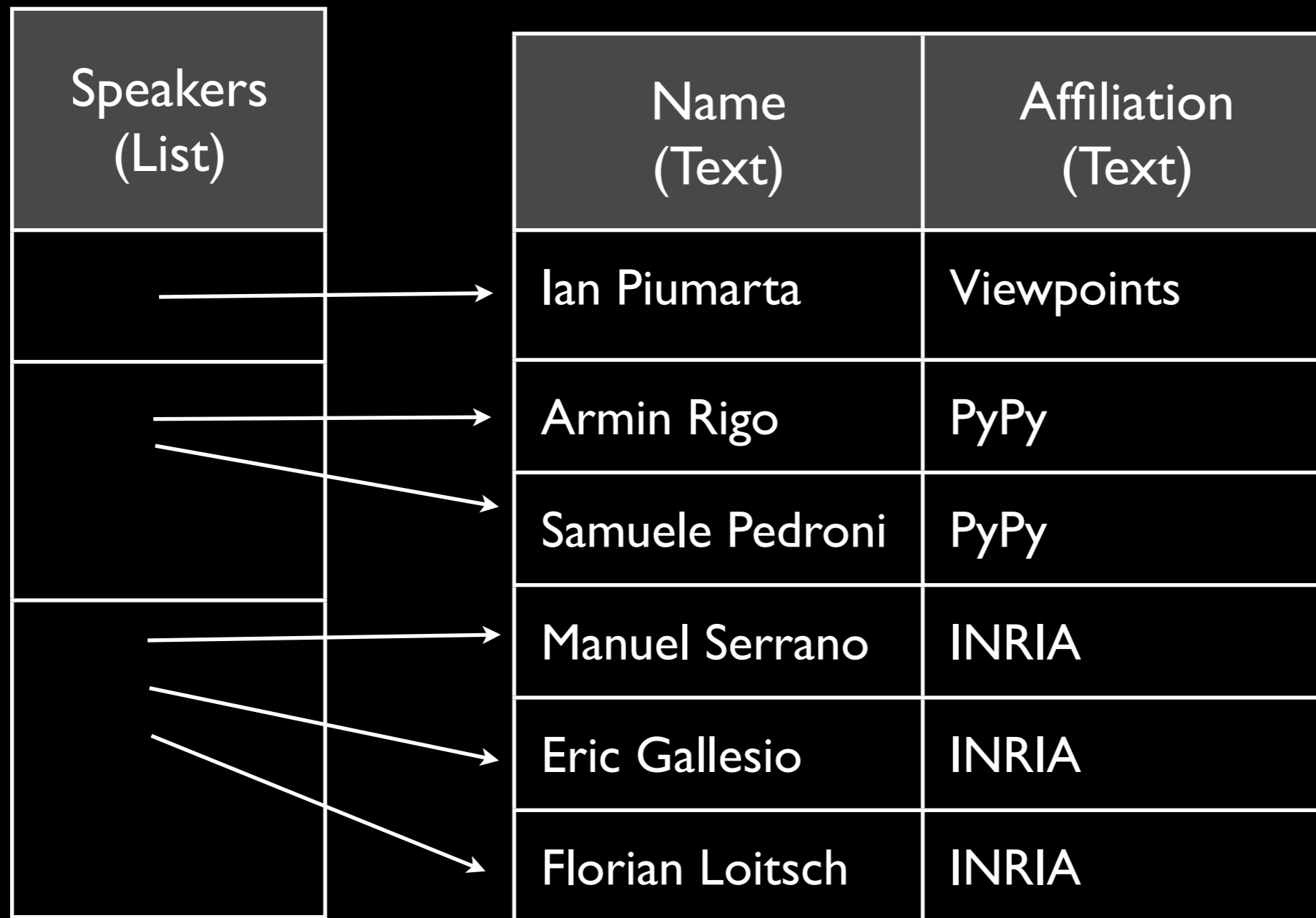
You have a field in a table which is more closely associated with a parent or child of that record.

Move the field from the original table to the related table. For each record in the original table, copy the field's value to the related record (or records) in the new table.

Move Field to Related Table



Move Field to Related Table



Merge Link with Link

You have two fields which model the same relationship with another table.

Merge the two fields into a new list field.

Also: Merge Link with List, Merge List with List

Merge Date with Date

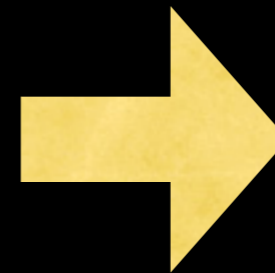
You have a date range whose start and end are defined by two separate fields.

Merge the two fields into a new date range field.

Also: Merge Date with Duration

Merge Date with Date

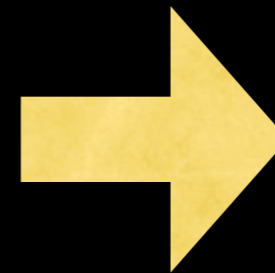
Start (Date/Time)	End (Date/Time)
Oct 23, 2006	Nov 6, 2006
Oct 23, 2006 6:00am	Oct 23, 2006 10:00am
Nov 4, 2005	Nov 5, 2005



Range (Date/Time)
Oct 23 - Nov 6, 2006
Oct 23, 2006 6:00am - 10:00am
Nov 4 - 5, 2005

Merge Date with Duration

Start (Date/Time)	Length (Duration)
Oct 23, 2006	2 weeks
Oct 23, 2006 6:00am	4 hours
Nov 4, 2005	2 days



Range (Date/Time)
Oct 23 - Nov 6, 2006
Oct 23, 2006 6:00am - 10:00am
Nov 4 - 5, 2005

Add Inverse Field

Oh, just demo it.