Why Simple Business Questions are not that simple … and what to do about it

Y. Cras
Chief Development Architect
SAP Business Objects
July 2011

Intention

BI is all about enabling business users to interact with the information that floods their organizations.

Business users want to ask *natural* business questions for which they expect an accurate, usable and *natural* answer.

After 20+ years, are we doing such a great job at it?

Agenda

- State of the Union
  - The historical promise of BI, and what people really took from it
- Simple questions, really?
  - Chasing sources of ambiguity
- Talking to the Liberal Arts Major inside
  - The love-hate relationship between BI and people
- What's In a Name?
  - The identity problem in BI
- Better Than a Thousand Words
  - The need for new metaphors

State of the Union

Did you realize that we've been dating for 20 years already?

Yeah, but honestly I'm still not quite sure if I want to commit to a long-term relationship.

BI System → ← BI User

Roots

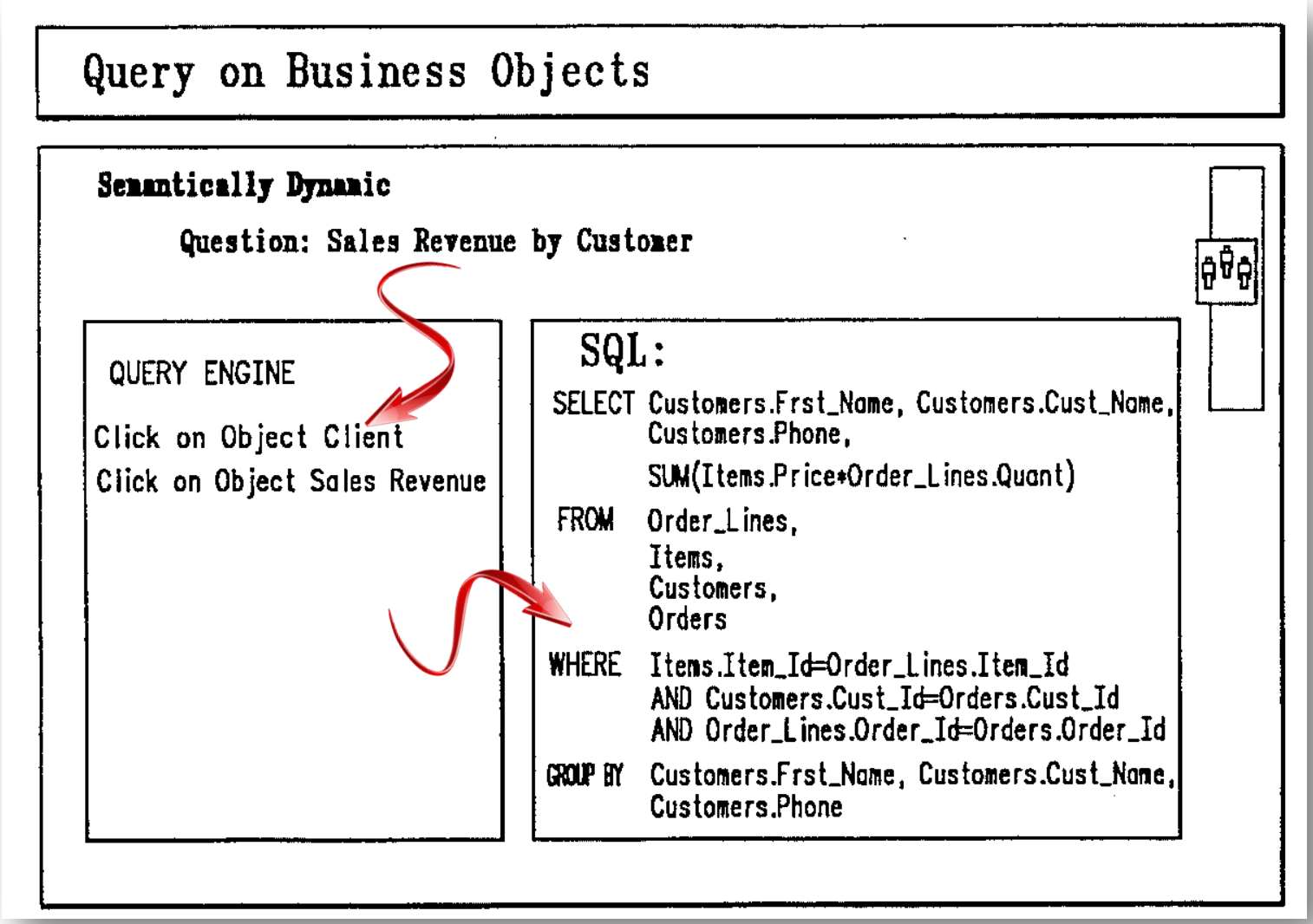
- US Patent 5,555,403 – Combot, Liautaud – Filed Nov. 27th, 1991

[57] **ABSTRACT**

A relational database access system and method provides a new data representation and a query technique which allows information system end users to access (query) relational databases without knowing the relational structure or the structure query language (SQL). The present invention utilizes semantically dynamic objects.

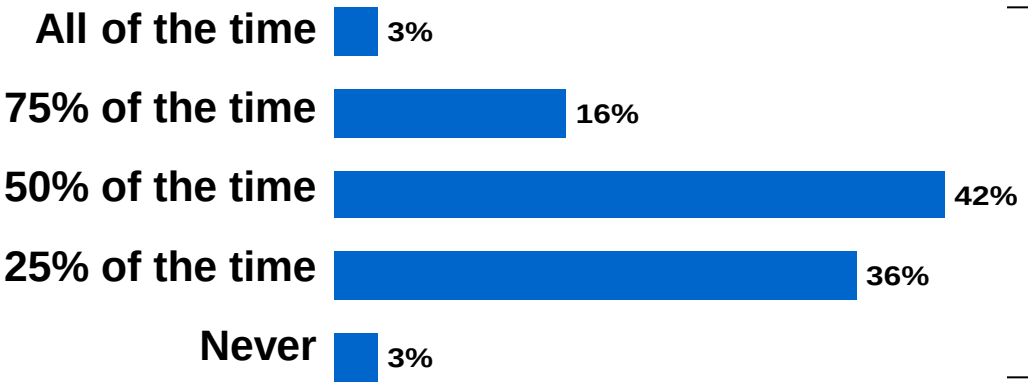
Semantically Dynamic Objects
The promise of Information democracy

- One of the best ideas after warm water. And it works, quite well at that.



Yet…

Gut Feel Decisions

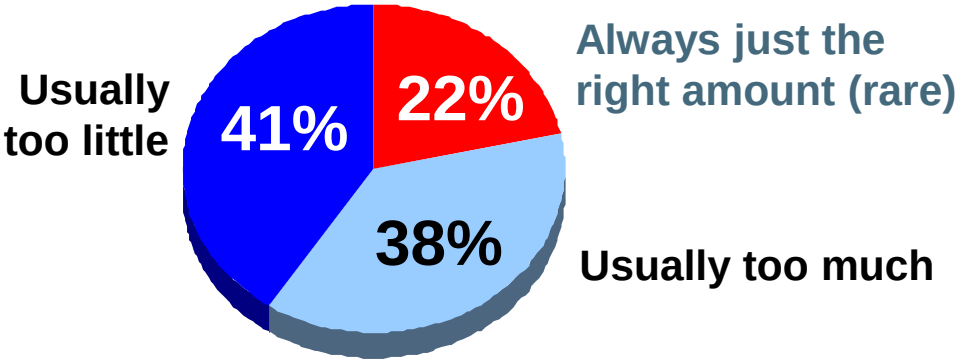


Base: 675 US and European business executives and managers
Source: BusinessWeek Research Services

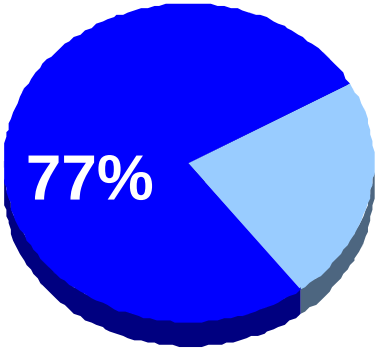
BusinessWeek

"Gut Feel" used by >60% of people, >50% of the time

Information available for important business decisions



Aware of bad decisions managers have made due to insufficient information



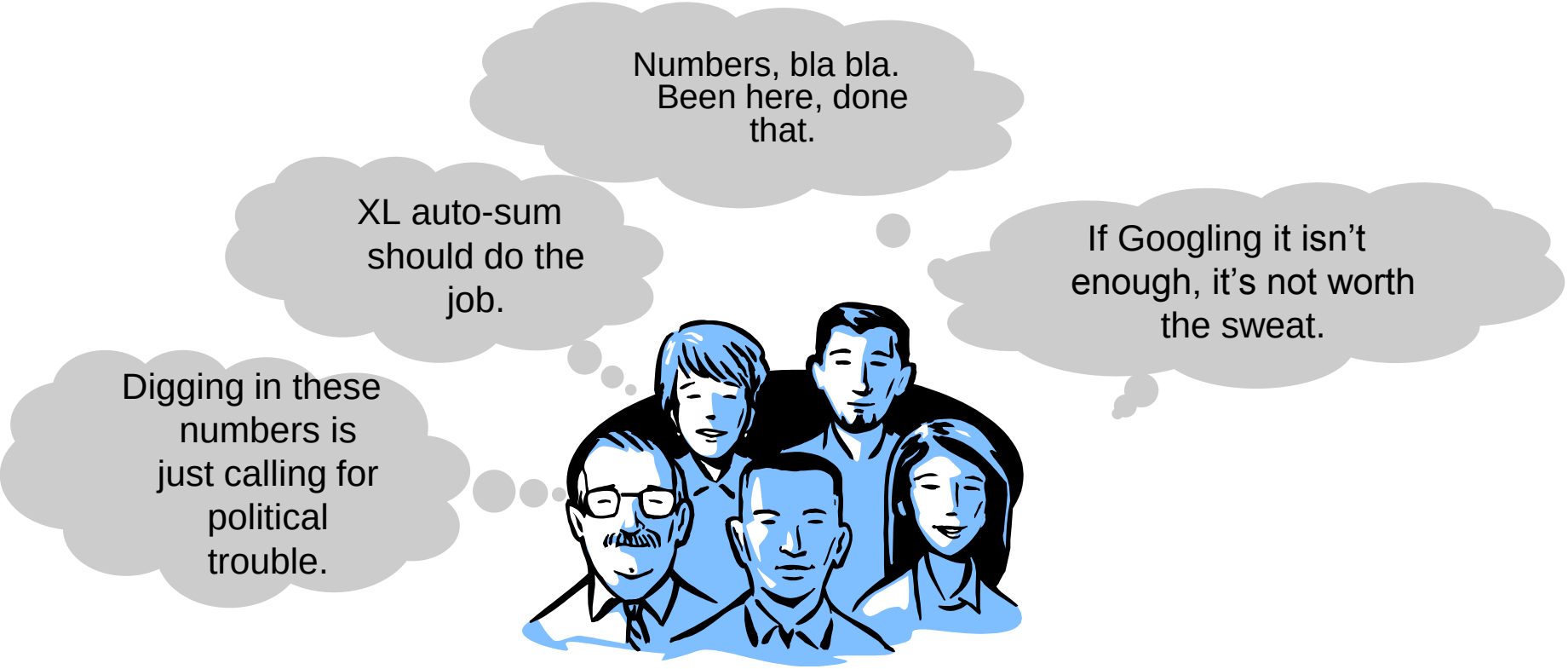
SAP BusinessObjects

Why is that?

The **Data** are there (quite a lot of it).

The Tools and **Technology** are there (to a large extent).

Most **People** are, well, left behind. Even for what we believe to be the simplest tasks.



What can we do about it:
Talk to our inner Liberal Arts Major.

Acknowledge that most people hate numbers and math.

Recognize that they need answers to business questions that are much harder to formulate than one would think.

And still, they and nobody else should be in power.

Don’t expect too much respect for structure from users. Give them the content they want.

Don’t expect complete and clean thought models. Accept and accompany incremental, trial-and-error approaches.

Gently guide them to what makes common sense (which also means out of Excel).

Make BI pervasive and invisible.

Simple Questions, Really?
Variations on a Theme

That's obvious. Well, wait…
Yes, that's obvious...
I think.

Sales Revenue By Customer

Query on Business Objects

Semantically Dynamic

Question: Sales Revenue by Customer

QUERY ENGINE

Click on Object Client

Click on Object Sales Revenue

SQL:

```
SELECT Customers.Frst_Name, Customers.Cust_Name,
       Customers.Phone,
       SUM(Items.Price*Order_Lines.Quant)
FROM   Order_Lines,
       Items,
       Customers,
       Orders
WHERE  Items.Item_Id=Order_Lines.Item_Id
       AND Customers.Cust_Id=Orders.Cust_Id
       AND Order_Lines.Order_Id=Orders.Order_Id
GROUP BY Customers.Frst_Name, Customers.Cust_Name,
       Customers.Phone
```


Sales Revenue for 2010 By Customer

Customer	Revenue
Arai	[Null]
Baker	162566
Brendt	[Null]
Diemers	[Null]
Dupont	[Null]
Durnstein	[Null]
Edwards	[Null]
Gentil	[Null]
Goldschmidt	18715
Hopkins	[Null]
Jones	[Null]
Kamata	112982
Kamimura	[Null]
Keegan	[Null]
Larson	101335
Makino	89085
Marlow	[Null]
Martin	[Null]
McCarthy	136989
McCartney	[Null]
Michaud	[Null]
Mukumoto	16720
Okumura	[Null]
Oneda	122906
Piaget	[Null]
Reinman	[Null]

Did you mean, *all* customers ?

… or just those who actually *bought something* that year?

Customer	Revenue
McCarthy	136989
Larson	101335
Baker	162566
Schiller	131282
Weimar	76730
Titzman	126090
Kamata	112982
Makino	89085
Oneda	122906
Goldschmidt	18715
Schultz	20330
Mukumoto	16720

You don't care? Well then, guess what happens next time you ask me for the *average* sales!

And, no, users won't tell an inner join from a LOJ.

Sales Revenue By Country

… for a travel agency. People from Country X fly to country Y.

Country	Revenue
Australia	[Null]
France	[Null]
Germany	1054248
Holland	[Null]
Japan	1046019
UK	[Null]
US	1186257

Did you mean, by country of *customers*?

… or by country of *destination* ?

Country	Revenue
Australia	[Null]
France	835420
Germany	[Null]
Holland	[Null]
Japan	[Null]
UK	[Null]
US	2451104

BI and analysis models expose different objects for each.

But then, how can you know which customers stay in their own country?

Country	[Local Revenue]
Australia	[Null]
France	[Null]
Germany	[Null]
Holland	[Null]
Japan	[Null]
UK	[Null]
US	888522

Destinations per Customer

Customer	Resort
Arai	Bahamas Beach
Baker	Bahamas Beach
	French Riviera
	Hawaiian Club
Brendt	Bahamas Beach
Diemers	Bahamas Beach
Dupont	[Null]

Did you mean, destinations where customers fly ?

… or destinations about which they ordered a guide?

Customer	Resort
Arai	[Null]
Baker	Bahamas Beach
Brendt	[Null]
Diemers	[Null]
Dupont	Bahamas Beach

BI Different Contexts or Cubes are used to isolate different facts. Not an easy job for a business user.

Correlations would also be nice for upselling.

SAP BusinessObjects

GNP per City

City	GNP
Augsburg	1054248
Berlin	1054248
Boston	1186257
Chicago	1186257
Cologne	1054248
Dallas	1186257
Kobe	1046019
Tokyo	1046019

Really? I can do that. Does not make much sense does it? Misleading even. But hey, you're the boss.

City	[GNP per Person]
Augsburg	1974,98
Berlin	162,21
Boston	228,16
Chicago	2207,97
Cologne	54,88
Dallas	2004,50
Kobe	59,36
Tokyo	3264,60

Prorated by city population it would mean something. But who am I to say that?

Some calculations don't make business sense.

But how to flag them and suggest smarter moves?

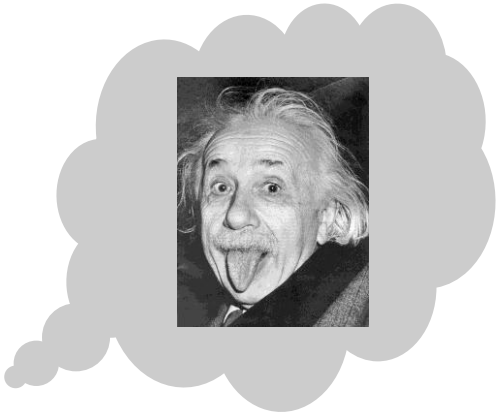
Sales Revenue by Customer (coda)

The semantics of a seemingly simple business question is strongly influenced by subtle yet critical details of how do determine an object from the others.

BI front-end must clarify the semantic links and help users navigate and understand them smoothly and intuitively.

Talking to the inner Liberal Arts Major

$$E=MC^2$$



WARNING: THIS COMIC OCCASIONALLY CONTAINS STRONG LANGUAGE (WHICH MAY BE UNSUITABLE FOR CHILDREN), UNUSUAL HUMOR (WHICH MAY BE UNSUITABLE FOR ADULTS), AND ADVANCED MATHEMATICS (WHICH MAY BE UNSUITABLE FOR LIBERAL-ARTS MAJORS).

Concept stolen from http://xkcd.com/

There's good AND bad in Human logic

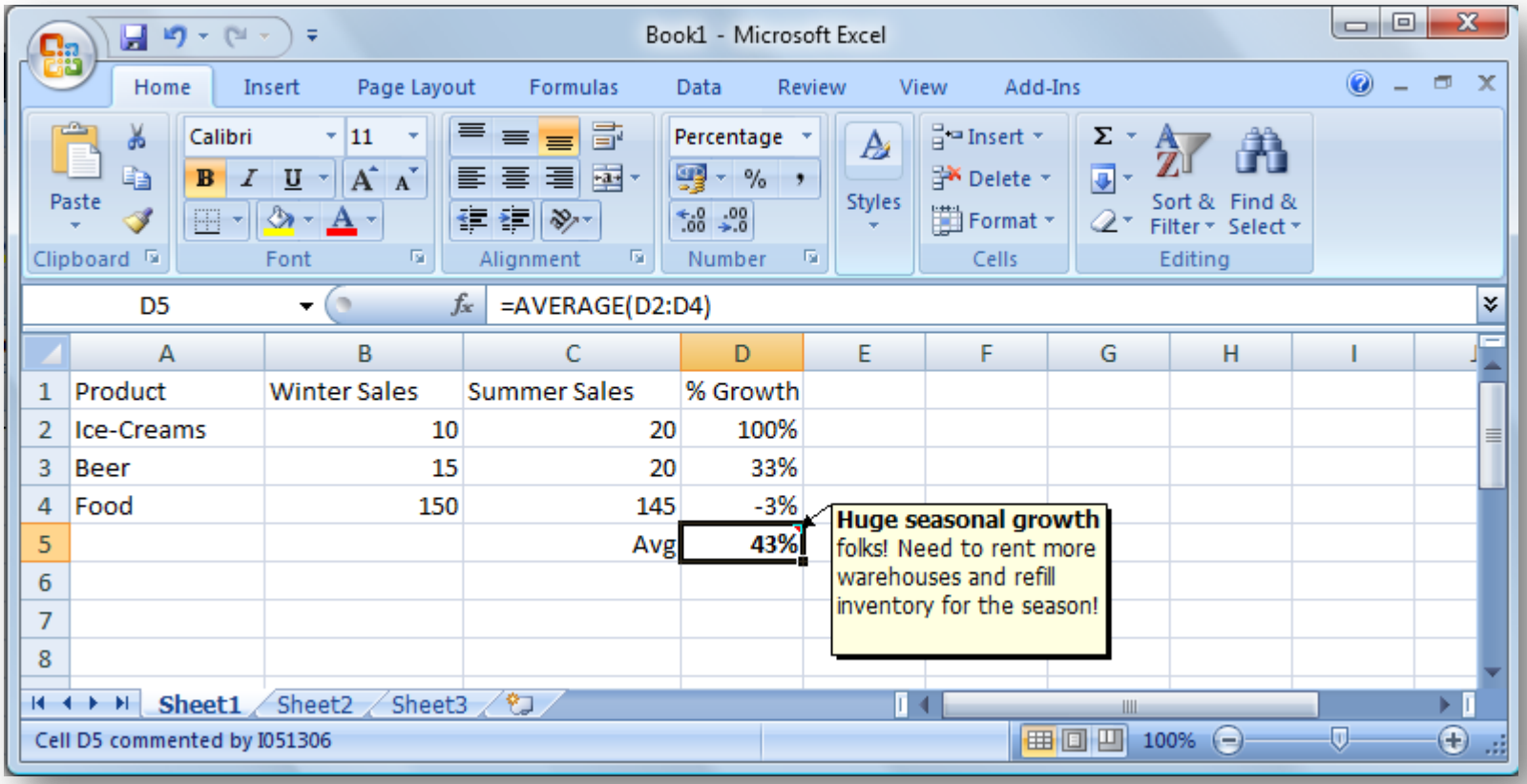
What sales did we do in France and Germany last year?

```
0 Rows
retrieved.
```

Most users don't know much about Boolean Logic, and (or?) care even less.

It may be too bad, but who pays our rents?

It looked god in Excel, though?



Most users will just do the usual things (for which Excel has optimized gestures). If nobody tells them there's something wrong, why would they worry?

What you'll never read in a novel (*good for you*)

What's most likely for the next Pulitzer award: this?

"Let me explain", he said, "I'm trying to find the sales generated by those customers aged 18 or less, living in the US or Caribbean, whose percentage growth for sales over last year over the 3 top-seller products in their geography was 10% greater than the average."

- "How fascinating", she replied.

-… or this?

- There are those customers, you know… the young ones, 18 or less.
- - Yes?
- - Well just take those who live, say, in the US. Add the Caribbean while you're at it, that's the same story.
- - OK fine. And?
- - Well now take, say, the 3 top-sellers products in each region. The real hits, you see?
- - I guess so.
- - Now look for those kids who bought more of that stuff this year than last year.
- - Quite a lot of them did.
- - Yep, let's prune that list a little. What's the average growth?
- - I'd say 5%.
- - Take only people much higher that that. Let's say, 10% more than average at least. We're talking fashion victims here.
- - Got it. That's still quite a list!
- - It's manageable. See how much money we did with those kids this year? That's our target audience.
- - You're a freaking genius!

Users don't think in subordinates, nested clauses, intricate filters. They go step by step, through trial and errors. We think declarative, they act incremental. We may be right, but again who pays our bills?

Depth-First vs. post-its on the fridge

```
Cake Recipe:
  Cook Cream.
  Bake pie.

Cook Cream:
  Boil Milk.
  Pour Eggs.

*** Null Egg Pointer ***

*** Cake Abort ***
```

I need to finish that cake. Where was I already? Oh, the cream. Shoot, I'm out of eggs. Wait, I'll do the baking first. Maybe just replace the cream with marmalade.

Most users are not too nervous about referential integrity or dependency management. They sort it out on the way, somehow, in no special order. They're OK with (non-linear) todo *lists*, certainly not with todo *trees.*

They may be wrong, but …

If I can't Google it, it's not worth the sweat.

2011 Sales for France please.

```
Did you mean:
Value of the measure Sales for the
Country France and the Year 2001, as
described in our sales datamart?
```

Are you stupid or what?

Users will use individual values and assume that the system infers which entity they represent, what relationships should be used between them, and even where the data for those lies.

They will specify structural entities (Customer, Year) only if they mean to query on them.

A Love-Hate Relationship

Well-designed BI	People
Has precise, consistent, sound logic and semantics	Navigate semantic halos
Does what it's told to do, doesn't double-guess nor nag the user	Make mistakes out of habit, of repetition
Relies on declarative, complete models	Favor experimentation, increments, re-use
Use dependency graphs and make sure references are resolved	Take notes on things to do and take care of them later if ever.
Uses fully qualified names	Live in ambiguity resolved by context.
Rely on complex structures	Infer structure from shape and proximity

How do you get along with your BI soft these days?

As they say on Facebook, "it's complicated".

What's in a Name?
The identity problem in BI

We have 100 headcounts

(working on projects)

Product.xlsx

We have 105 headcounts

(in the org chart)

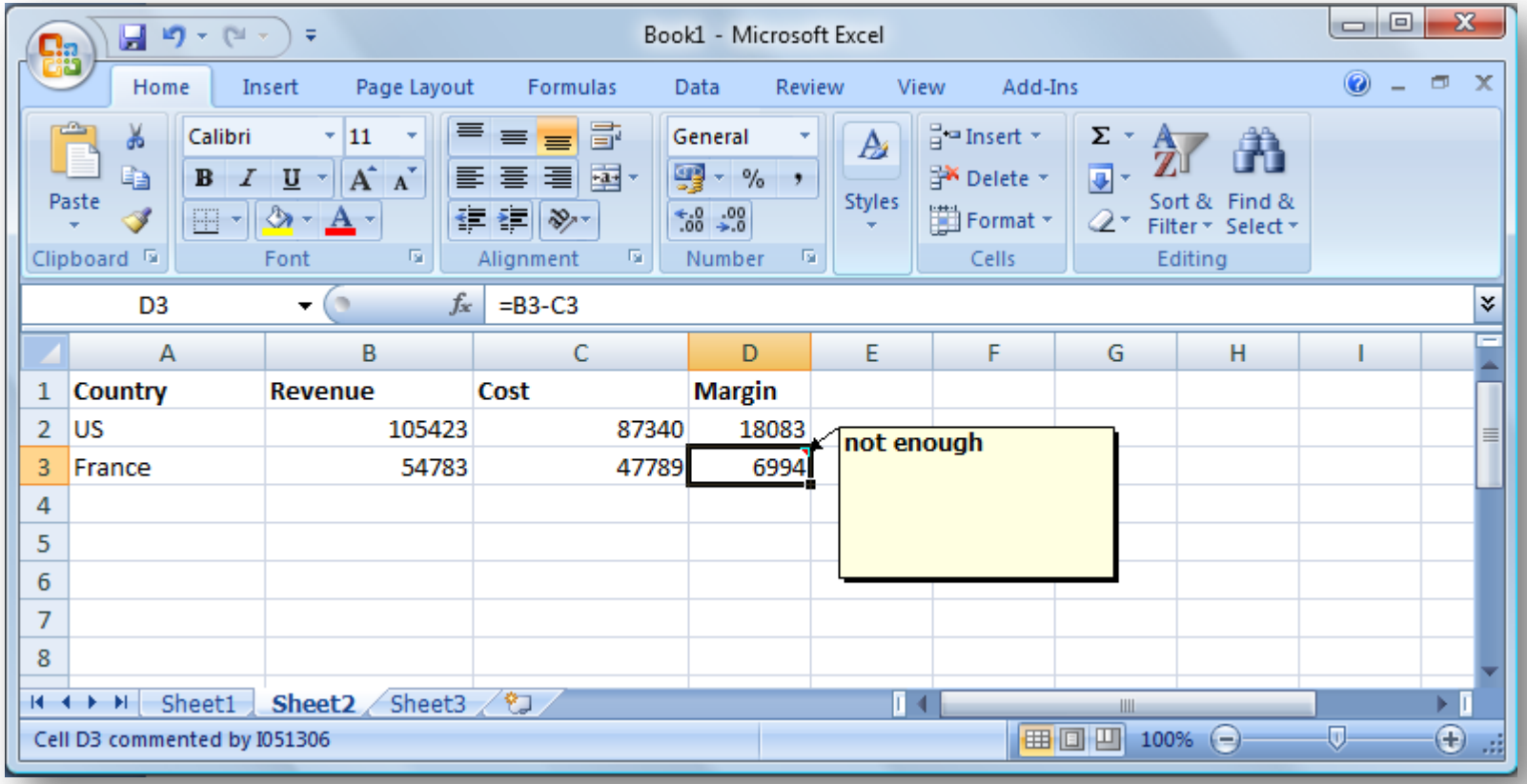
HR.xlsx

We have 107 headcounts

(on the payroll))

Finance.xlsx

Why Spreadsheets Suck
(well, one of the many reasons why)



They refer to cells, not content.

They annotate cells, not content.

What is “not enough”? The margin for France (this quarter)? Or B3-C3?

1st Prize: a unique ID.
2nd Prize: *two* unique IDs.

Corporations have a big problem: identity uniqueness.

- How do I know that this customer in your books is the same customer in mine?
- How do I know that your catalog reference is that product ID?
- Master data management is a huge pain point.
- And that's in a corporation that can *force* you to comply.

The Semantic Web has the same problem

- See sameAs.org.
- 4 URIs (last time I checked) for the city of London
- And that's only between people who know each other and want to collaborate.

BI's problem is even worse. BI is a lot about data that exists only in *intention*, that will be materialized only (if ever) in a *transient* way.

- What's the URI of "the revenue on July 3rd, 2011 of product (then named as) PX2211"?
- What defines the standard corporate way of computing KPEX amortization? Is it a file? A piece of data? A mathematical formula?

The fridge works but the door is open

The IT industry spends more money integrating systems and repairing information leaks than creating new value.

We don't (or rarely) lose data.

We constantly lose and re-invent meaning. Hence redundant design, multiple sources of truth, entropy.

Either we do low-cost BI with uncertain semantics. Or high-TCO enterprise BI.

Can't we find more creative ways of preserving meaning just as we preserve content?

The Challenge of Identifying BI Data

Identify the Building Blocks of BI expressions

- Dimensions and measures of Universes or Cube Models
- URI of standard linked data

Unambiguously identify calculations

- "Revenue – Cost" should mean the same thing in all contexts
- Neither ontologies nor ER or UML capture aggegation.

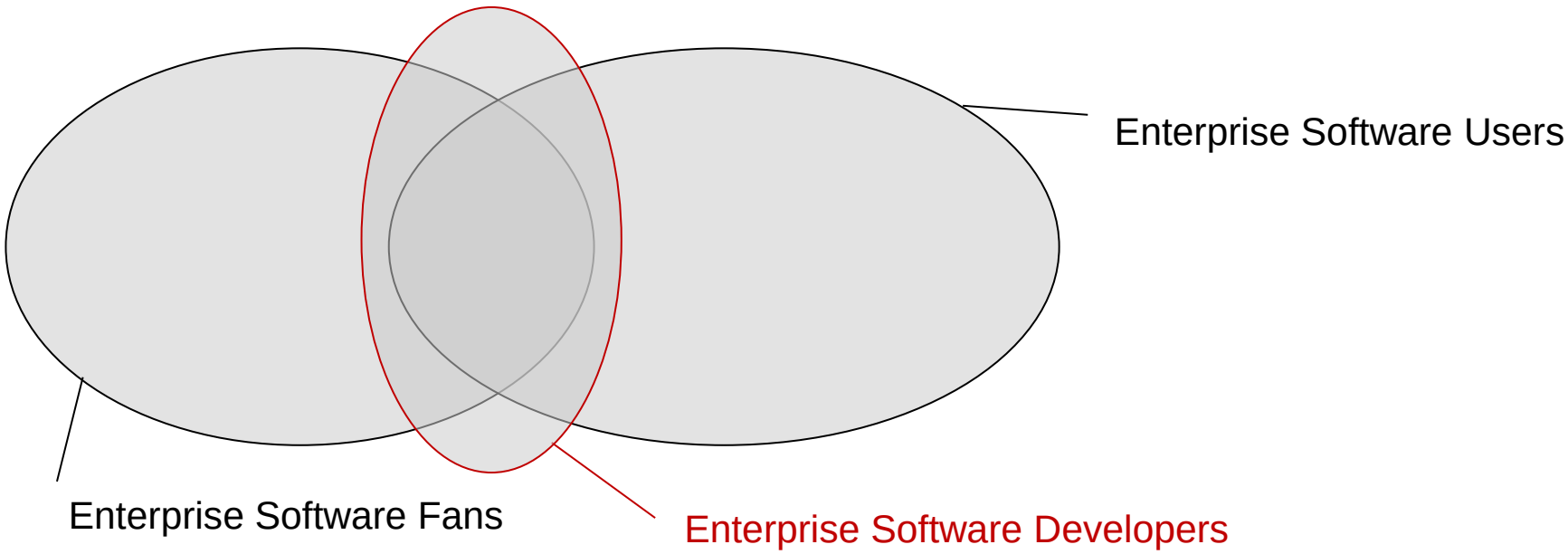
Completely specify the dimensional context.

- "the revenue on July 3rd, 2011 of product (then named as) PX2211"
- For which set of customers? Which version of the product catalog?
- Also a legal/compliance issue

Index and refer to the data *definition*

- Not only its value
- Not only the document that contains it.
- Semantic Search, search to query, text to query
- Declarative, shareable, self-sufficient BI definitions

Better Than a Thousand Words



Concept stolen from http://thisisindexed.com/”

The need for new metaphors

We should not give up on good qualities of BI software and models.

But we need to find the right metaphors to expose them to people.

We have some of them.

More are to be invented.

Boolean Logic is sometimes intuitive

Sales | Paris | 2001

Sales in Paris AND in 2001

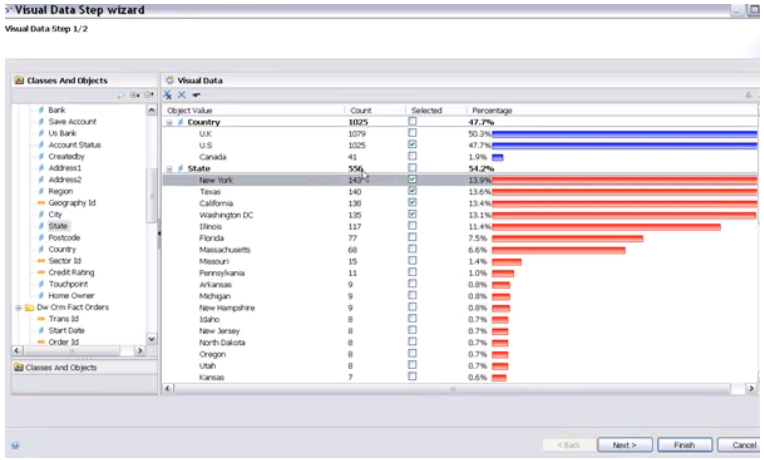
Different dimensions

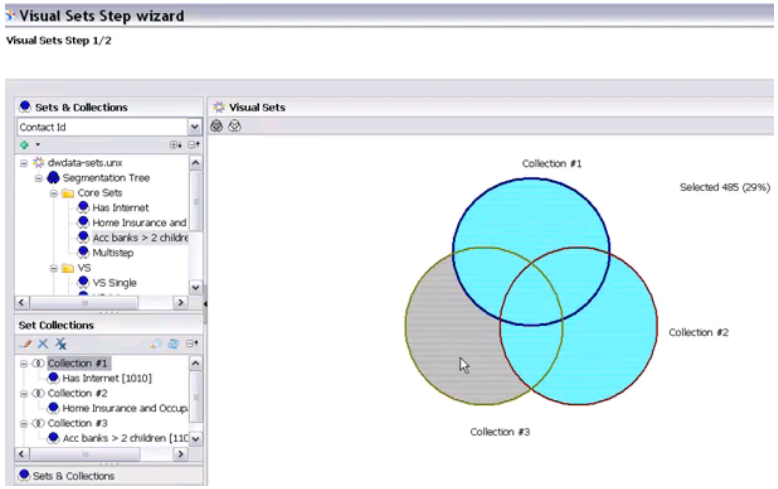
Sales | Paris
Lyon

Sales in Paris OR in Lyon

Same dimension

Visual, Incremental Set Building



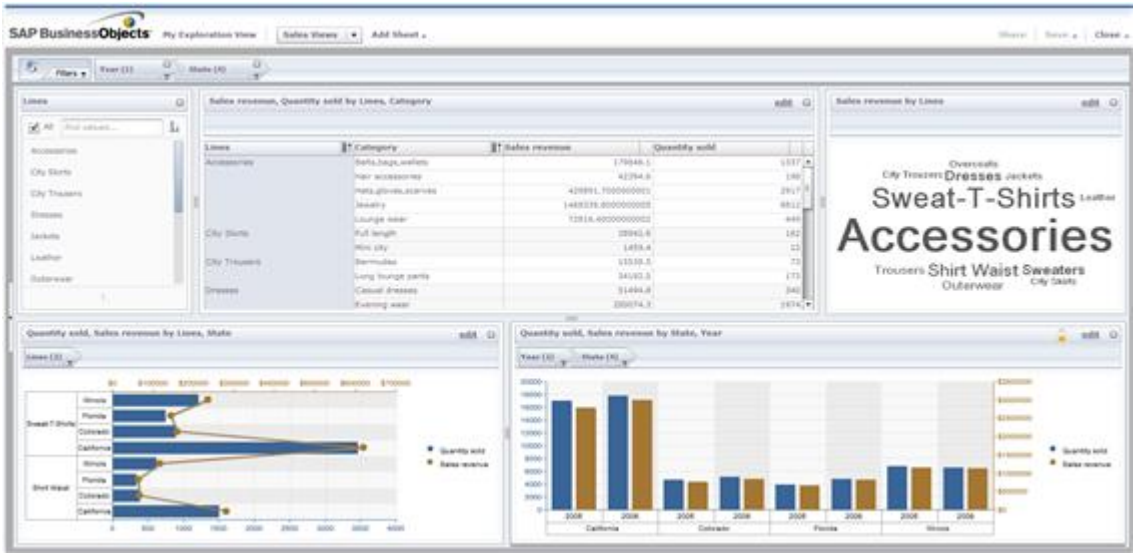


- See your data quickly
  - Fast intuitive interface
  - Explore large volumes with ease
- Understand the context
  - Use counts/percentages
  - Always select from lists
  - Intelligently limit choices to show only real values
- Help to define your questions
  - Learn from your data
  - Change direction as you learn
  - Select the right data easily

Source: G. OLIVA, SAP

SAP BusinessObjects

Semantic Visualization



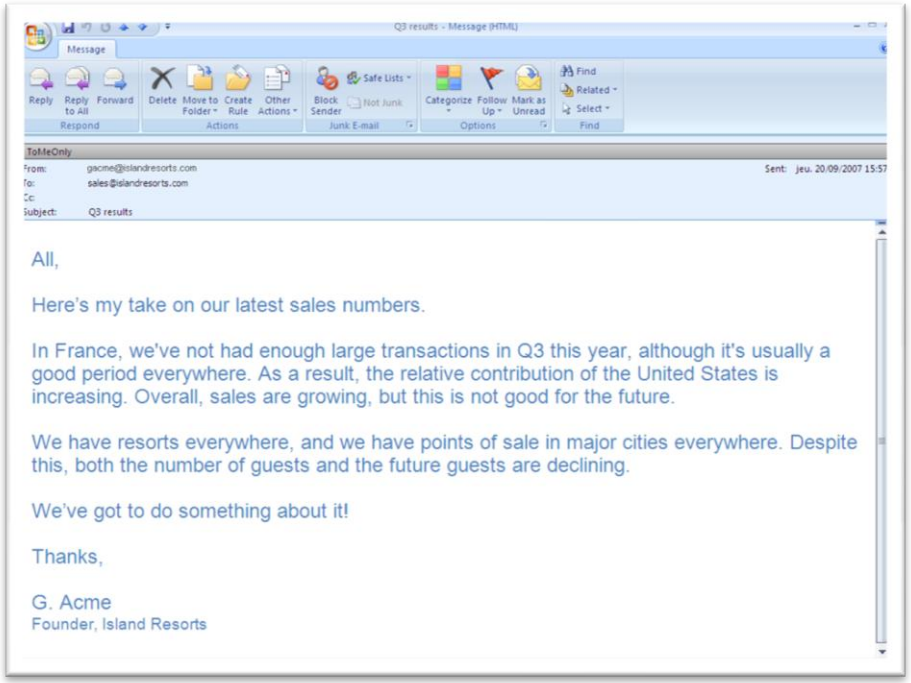
"Say it with Charts"

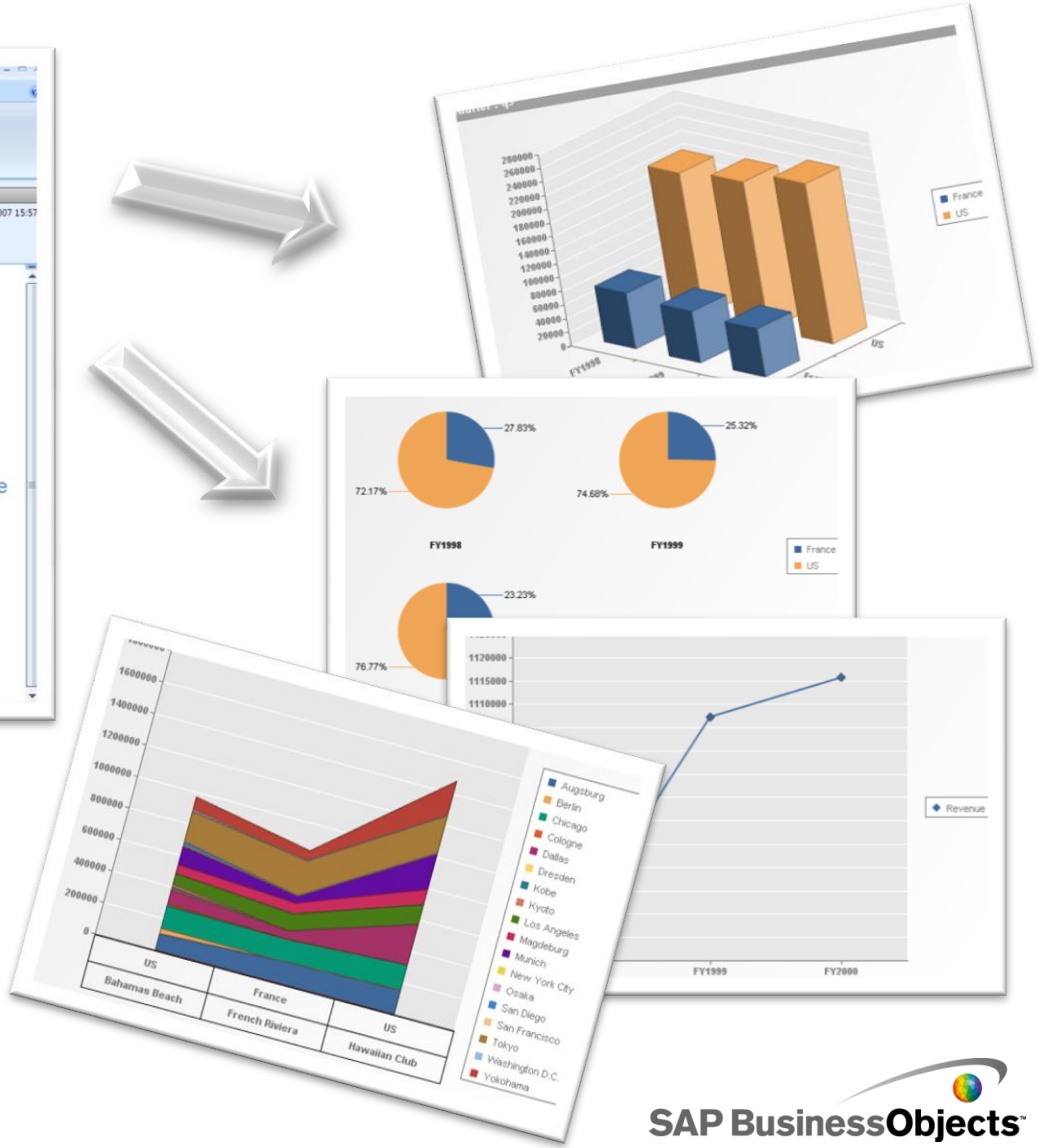
- Understand the business semantics of time, Geography, organizations
- Analyze the topology and content of a data set
- Suggest the best charts and dashboards to represent it, based on best practices

Gestures

http://vimeo.com/23899403

Inferring structure from content





Also, ontology inference

Conclusion:
We love BI best when it's nowhere to be seen

How do you get along with your BI soft these days?

What BI soft?

I have all the business data I need everyday.

They say the data fairy brings it when you're not looking.

How cute!

Some couples last longer when they don't meet too often…

Why Simple Business Questions are not that simple … and what to do about it

Y. Cras
Chief Development Architect
SAP Business Objects
July 2011