OpenOffice: is it getting better ?

February 2009

Michael Meeks

michael.meeks@novell.com

“Stand at the crossroads and look; ask for the ancient paths, ask where the good way is, and walk in it, and you will find rest for your souls...” - Jeremiah 6:16

Novell®

Overview

- What is OpenOffice.org – really
  - two views of what it should be
  - an exciting future direction ?
- What is new ?
  - fun new features & fixes, misc. demos, etc/
- Getting under the bonnet ('hood')
  - manufactured from the purest Duct-Tape ?
  - standards, file-formats etc.
- How to get involved
- Conclusions

Understanding OpenOffice.org

An un-authorised view ...

- OpenOffice.org
  - Community partially inspired by Freedows (?):
    - *"... the projected features, combined with sizable worldwide developer and alpha/beta-testing interest, resulted in difficulties in project organization and coordination. The project made little progress towards producing running code ..."*
  - Managed by volunteer manager **visionaries** ™
- ooo-build / go-oo.org – where my heart is.
  - Fun & hackerly, aims at being 'real' open-source project
  - relational, distributed, scalable, simple, minimal
  - sad to be in the 'beginners' track today: straw-poll - **hackers?**
  - distributed by ~all Linux Distros.
  - Excited by Free-software vs. *'professional S/W development'*

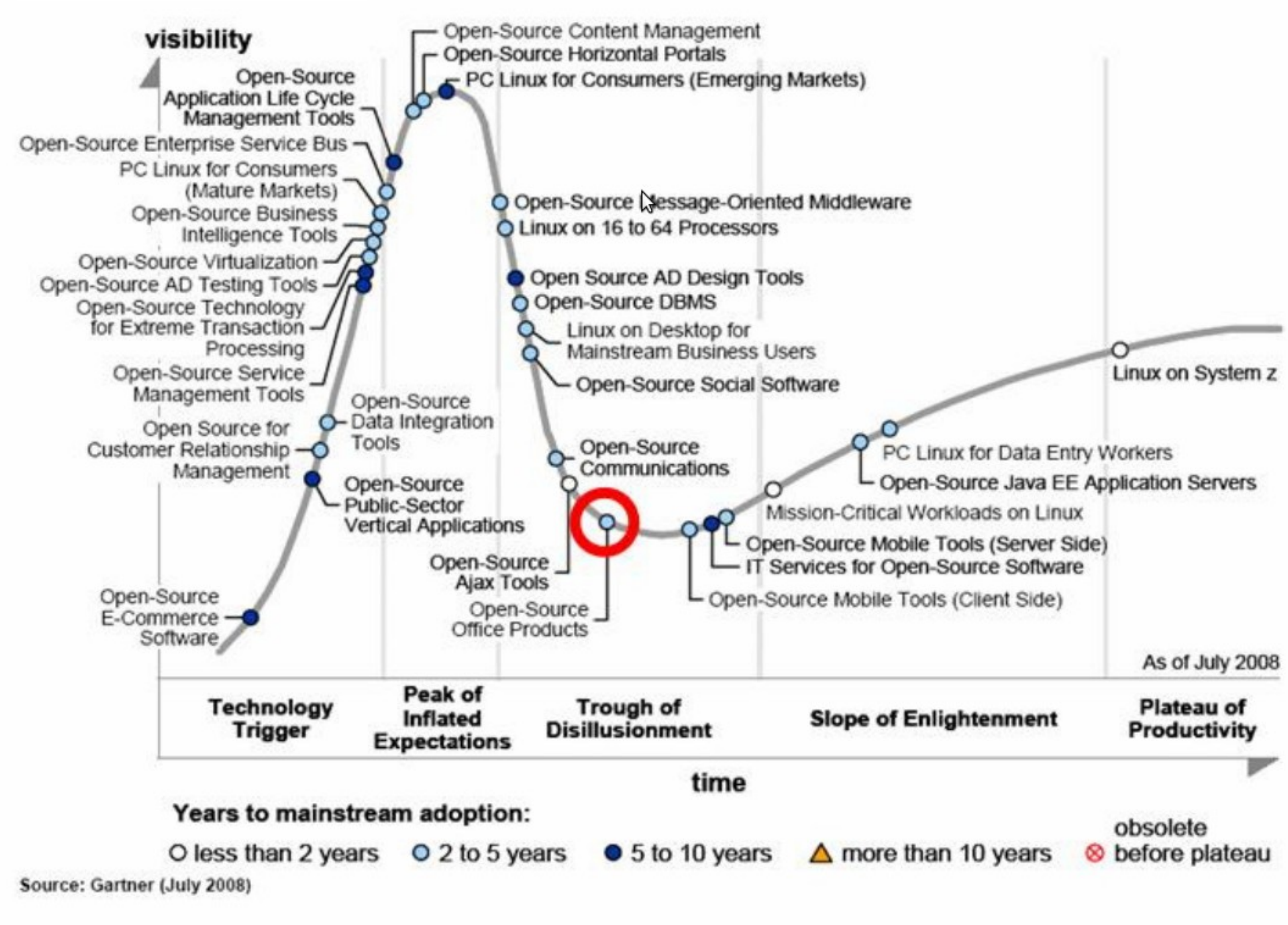
Tedious Politics

The trappings of success ...

- OpenOffice.org
  - 100 million users (ish)
  - a real alternative – people using it exclusively
  - tons of Government deployments, some business
  - 40 million downloads of Sun's OO.o 3 in 4 months
    - around 2 downloads / second since then, ~2+ million / week sustained
- Extensions
  - http://extensions.services.openoffice.org/
    - Where is that missing feature you really need, but now you're off-line ?
  - Templates, dictionaries, PDF import, Presentation viewer
    - Most popular 128k this week (French dict),
    - presentation viewer: 2.5k (0.1%)
  - go-oo builds in the ones that make sense … saves ABI grief.

Gartner's latest view:

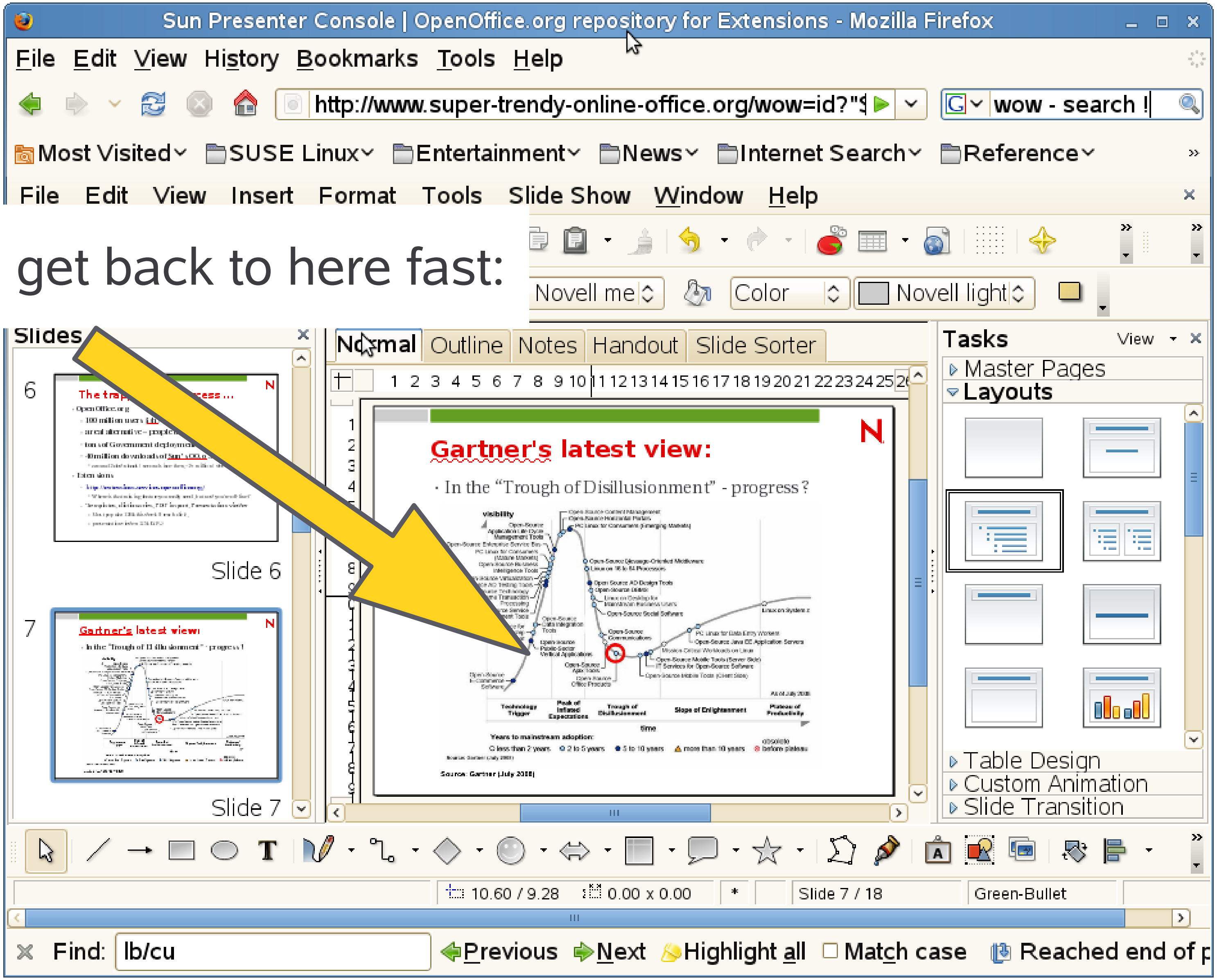
- In the “Trough of Disillusionment” - progress ?



Source: Gartner (July 2008)

An idea for success ! Web 2.0 !

Lets get back to here fast:

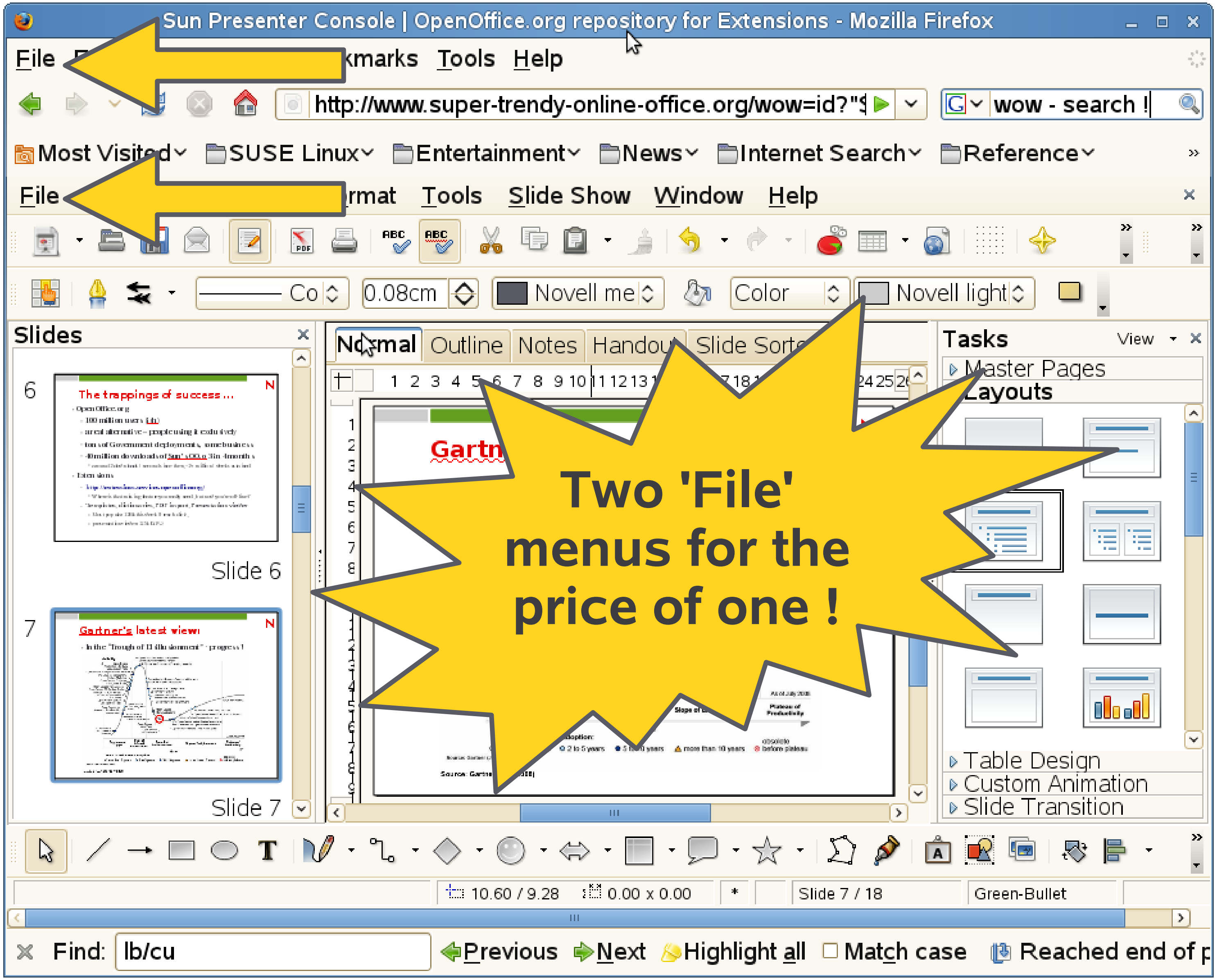

Sun Presenter Console | OpenOffice.org repository for Extensions - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://www.super-trendy-online-office.org/wow=id?"$
wow - search !
Most Visited SUSE Linux Entertainment News Internet Search Reference
File Edit View Insert Format Tools Slide Show Window Help
Novell me
Color
Novell light
Slides
6
Slide 6
7
Slide 7
Normal Outline Notes Handout Slide Sorter
Tasks View
Master Pages
Layouts
Table Design
Custom Animation
Slide Transition
10.60 / 9.28
0.00 x 0.00
*
Slide 7 / 18
Green-Bullet
Find: lb/cu
Previous Next Highlight all
Match case
Reached end of p

Gartner's latest view:

- In the "Trough of Disillusionment" - progress ?

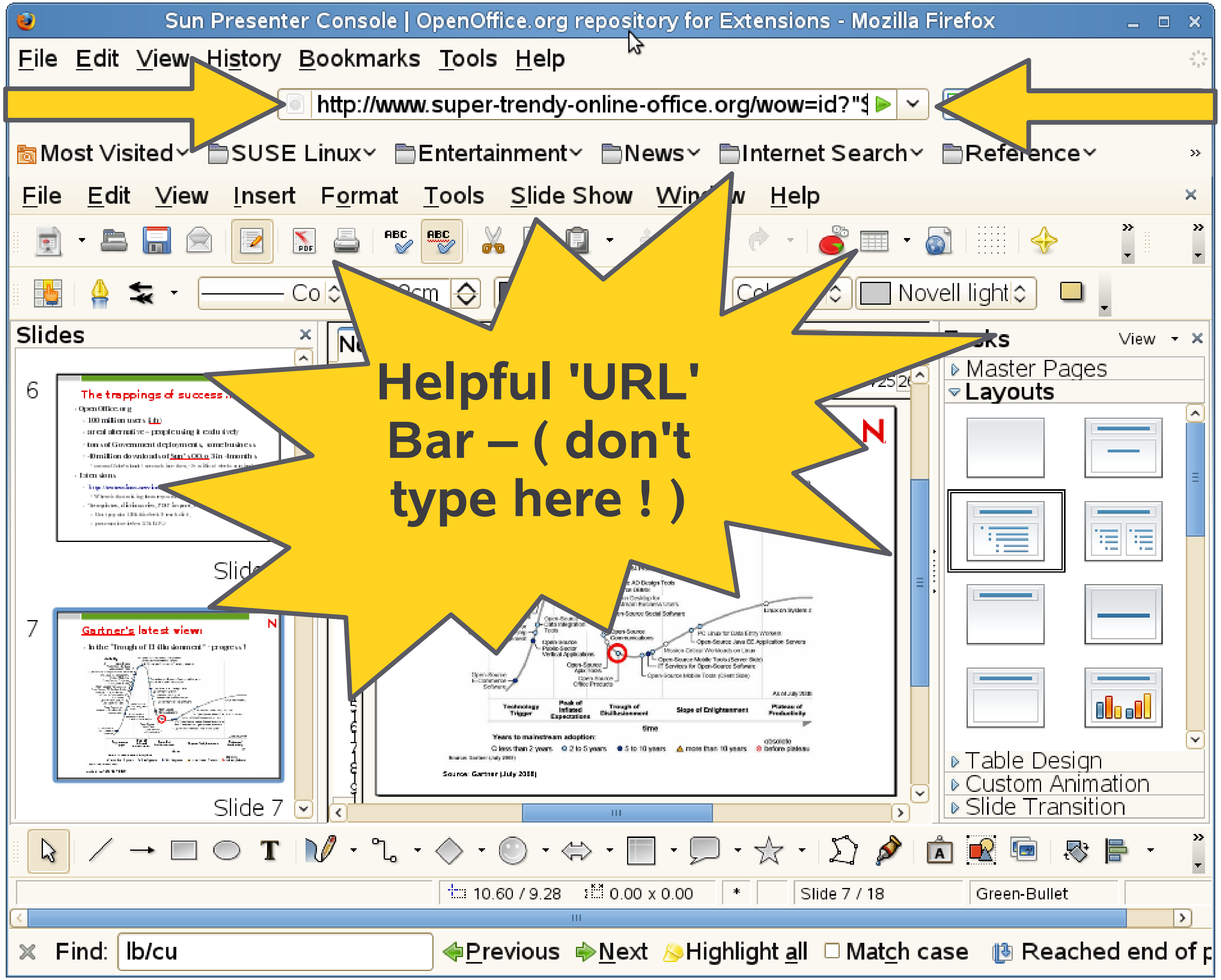
Source: Gartner (July 2008)

An idea for success ! Web 2.0 !



Two 'File' menus for the price of one !

An idea for success ! Web 2.0 !



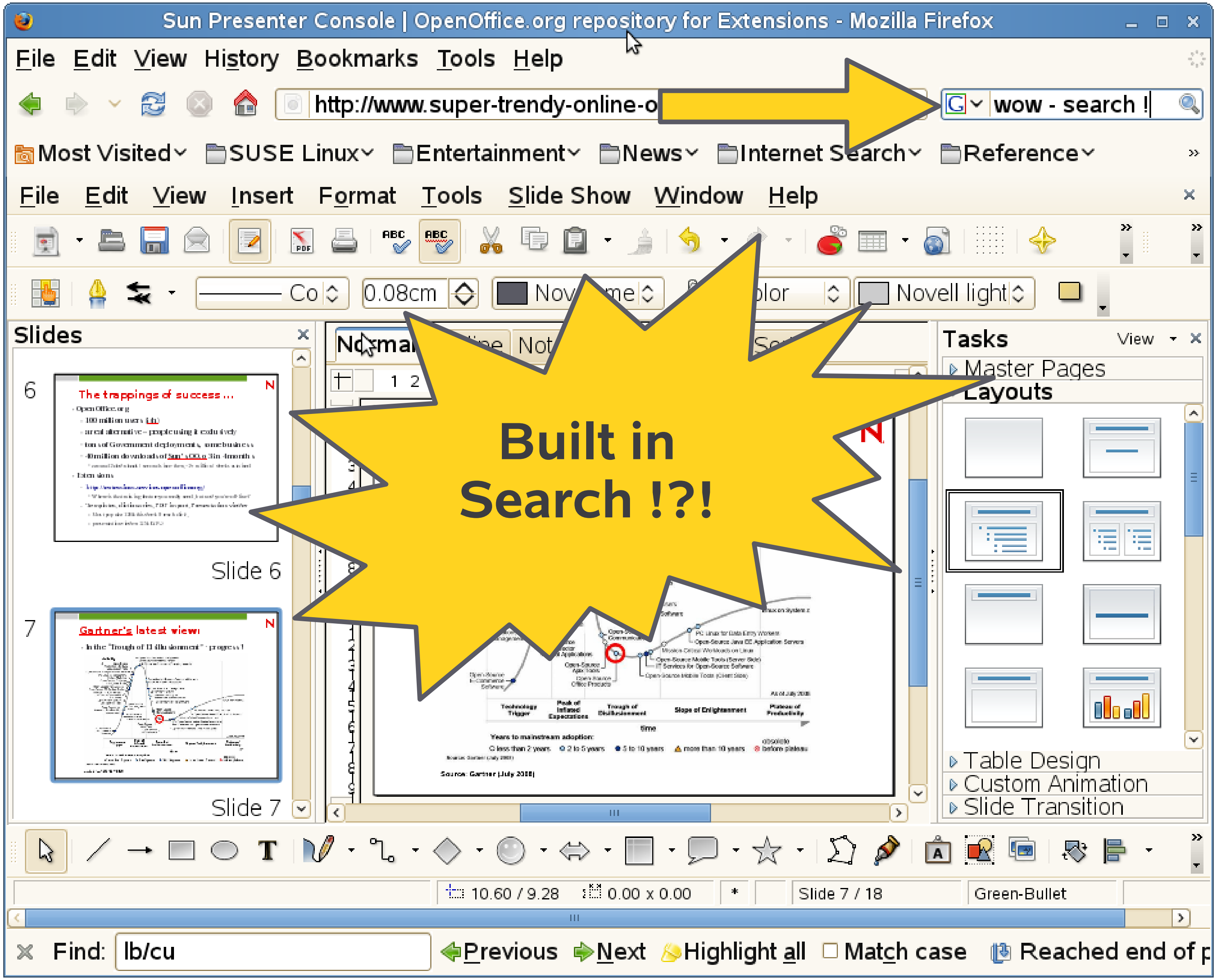
Sun Presenter Console | OpenOffice.org repository for Extensions - Mozilla Firefox

http://www.super-trendy-online-office.org/wow=id?"$

Helpful 'URL' Bar – (don't type here !)

Find: lb/cu

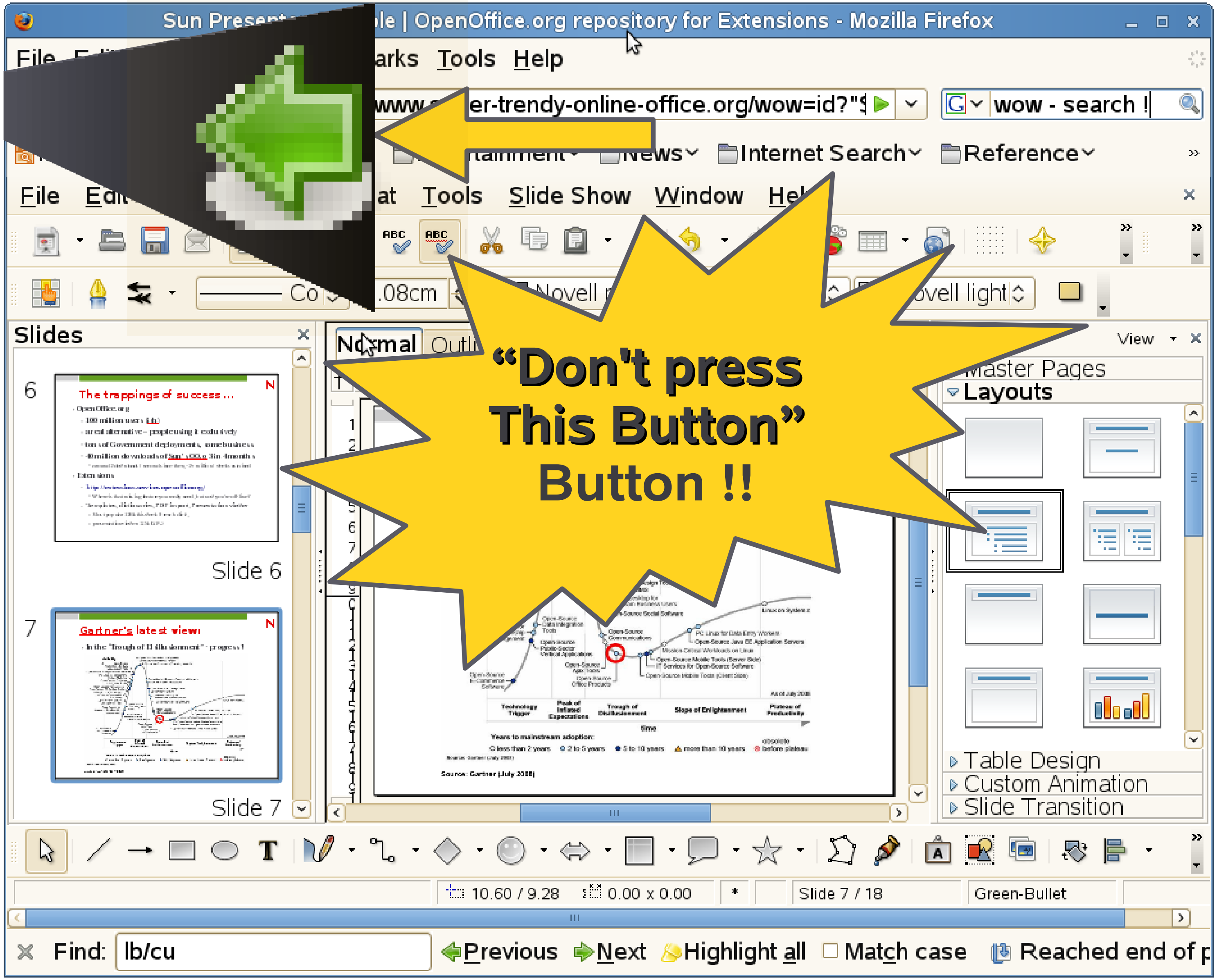
An idea for success ! Web 2.0 !



Built in Search !?!

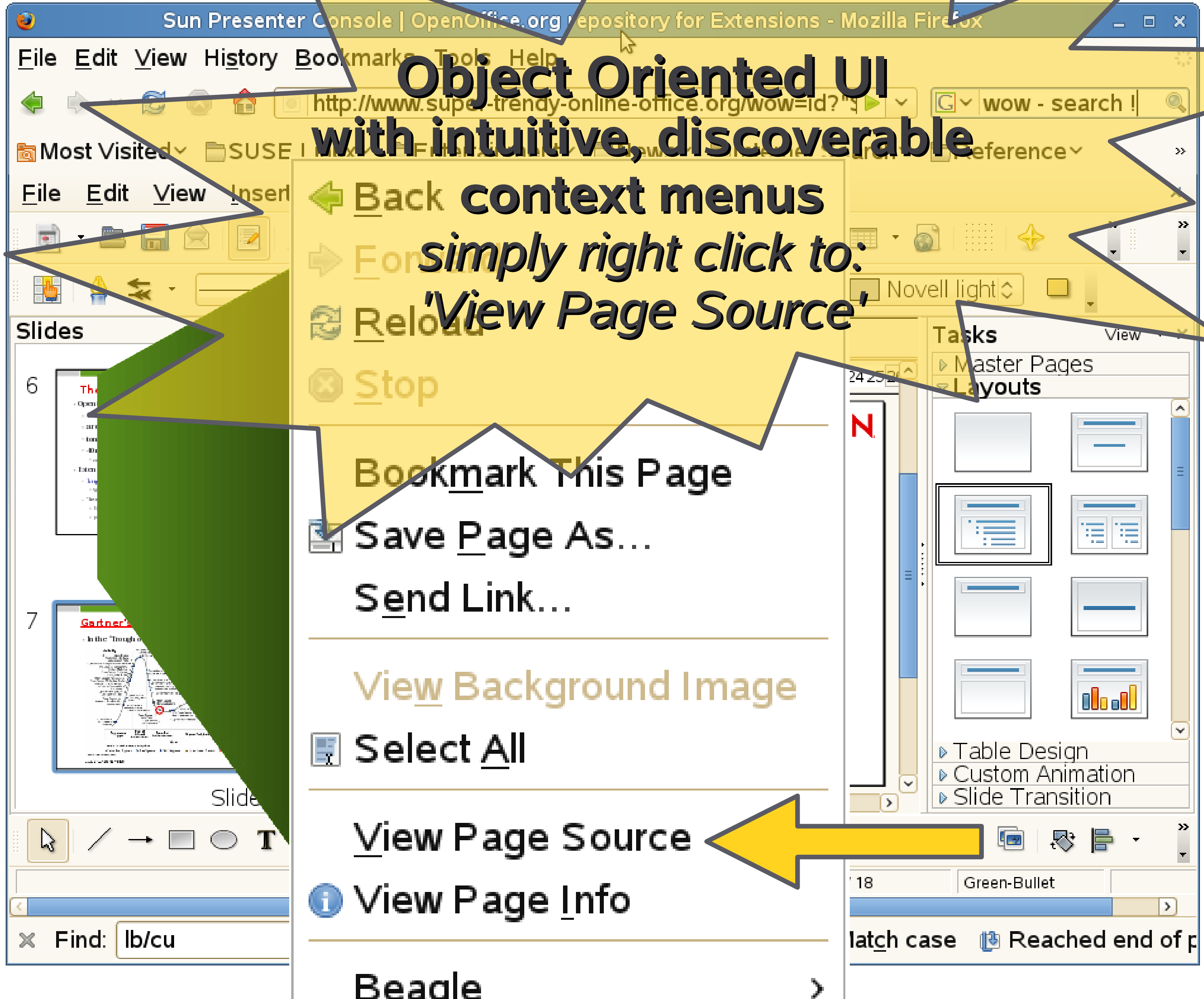
Sun Presenter Console | OpenOffice.org repository for Extensions - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://www.super-trendy-online-o
wow - search !
Most Visited SUSE Linux Entertainment News Internet Search Reference
File Edit View Insert Format Tools Slide Show Window Help
0.08cm
Novell light
Slides
Tasks View
Master Pages
Layouts
Table Design
Custom Animation
Slide Transition
Slide 6
Slide 7
The trappings of success ...
Gartner's latest view:
Source: Gartner (July 2008)
10.60 / 9.28
0.00 x 0.00
Slide 7 / 18
Green-Bullet
Find: lb/cu
Previous Next Highlight all Match case Reached end of p

An idea for success ! Web 2.0 !



“Don't press This Button” Button !!

An idea for success ! Web 2.0 !



Object Oriented UI with intuitive, discoverable context menus

simply right click to: 'View Page Source'

Other key features ...

- 100% compatible with all existing browsers*
  - Internet Explorer 8, Firefox 3.2 or Safari 3.2.3 required
- For off-line use requires:
  - Grinding Cogs extension version 2.17.3 – and 1Gb of memory.
  - A local copy of OpenOffice.org to do document import/export
- For reasonable performance:
  - FastJavaScript ™ extension is required, quad-core recommended, **8 million lines of JavaScript ...**
- “Tunnel” feature (*“Loading...”*)
  - Warns you before loosing what you just typed as you go through a tunnel.
- *“To display this page the application must send information that will repeat any action (such as a search or order confirmation that was performed earlier)”*

Returning from la-la land …

Some much less silly ideas:

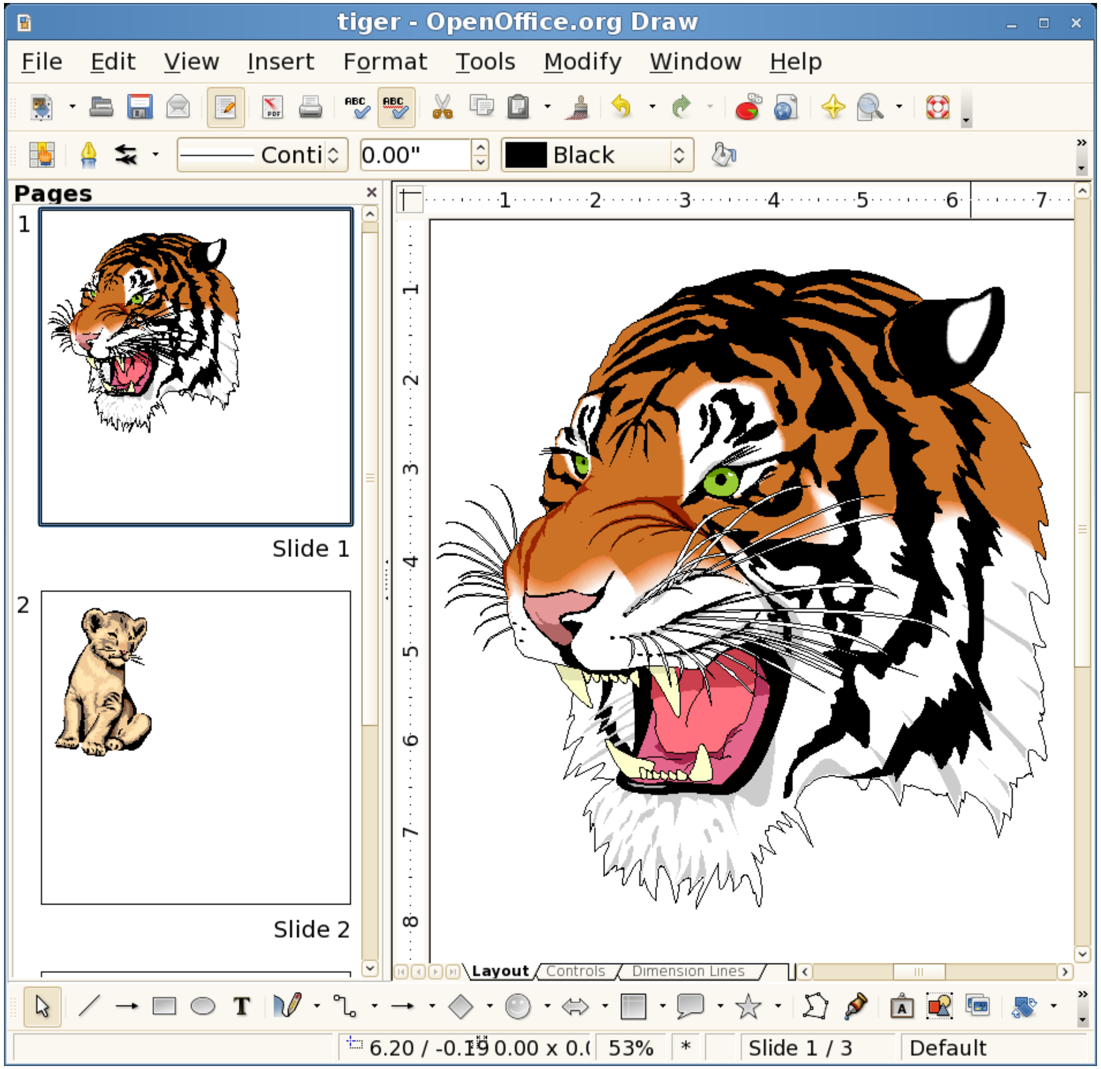
- ODF@WWW
  - an ODF Wiki – complete with rich editing
  - Kay Ramme's baby
    - http://odf-at-www.openoffice.org/
  - actually partially working
- Even better models: Live Editing ...
  - AbiCollab type thing:
  - “do whatever you like, but do it to all documents in the same order”
- Big hole in OO.o here currently ...

Fun new stuff ...

New things ...

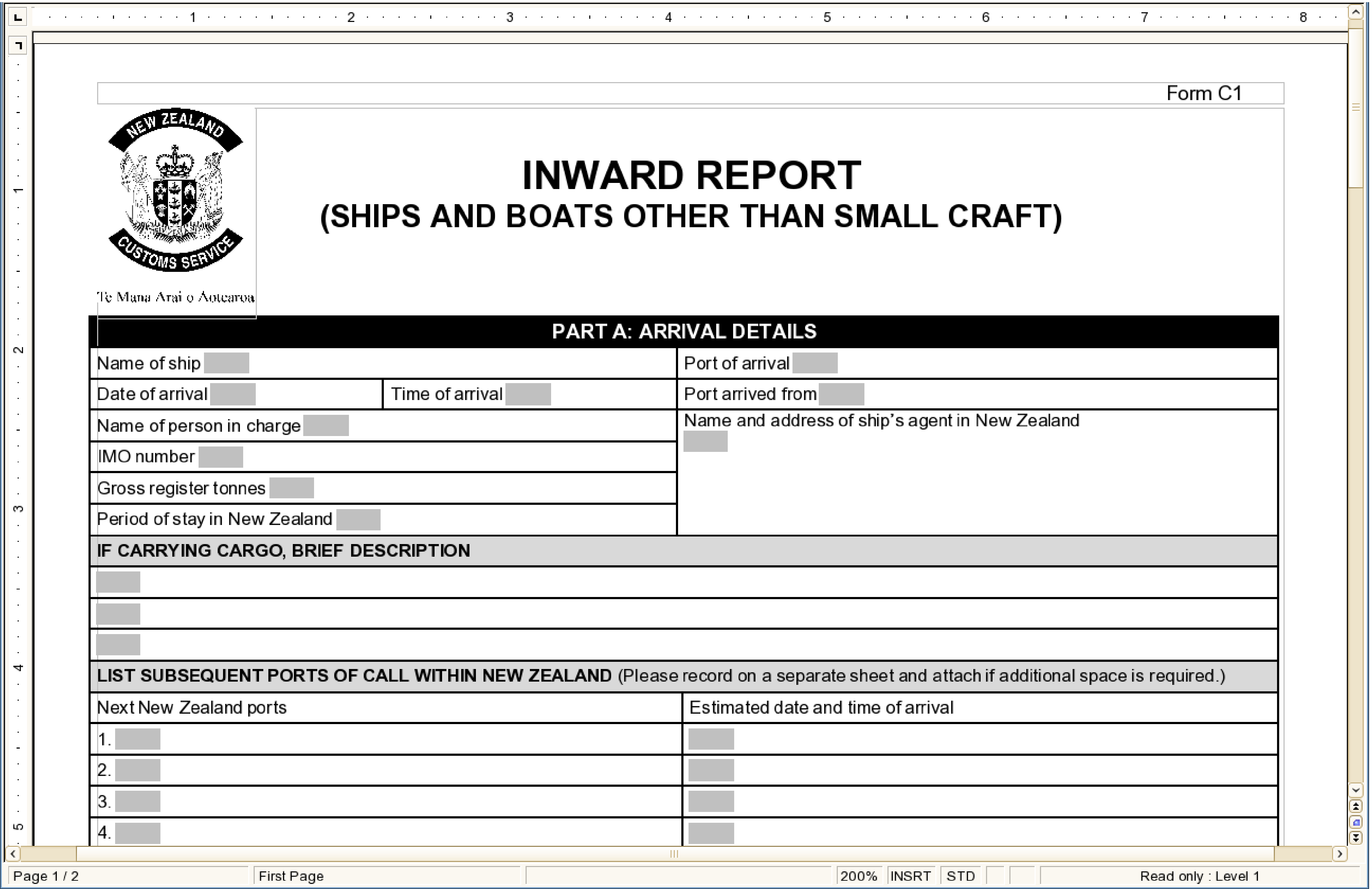
- OpenGL transitions (as you see)
  - Linux / OSX only … - ergo disabled in the up-stream builds.
- Command-line file format converter:
  - `ooconvert  foo-baa.doc  foo-baa.html`
    - Instant [or something] web view.
- Split build … build just 'Writer' in 15mins
- Space: splitting non-common file filters in calc / writer
- Time: finally including more work from go-oo
  - (no) symbolic link resolution
  - finally shrinking / improving .RDB file performance.

SVG import:



tiger - OpenOffice.org Draw
File Edit View Insert Format Tools Modify Window Help
Conti
0.00"
Black
Pages
1
Slide 1
2
Slide 2
Layout
Controls
Dimension Lines
6.20 / -0.19
0.00 x 0.0
53%
*
Slide 1 / 3
Default

Powerful & Interoperable fields:



Form C1

NEW ZEALAND CUSTOMS SERVICE

Te Mana Arai o Aotearoa

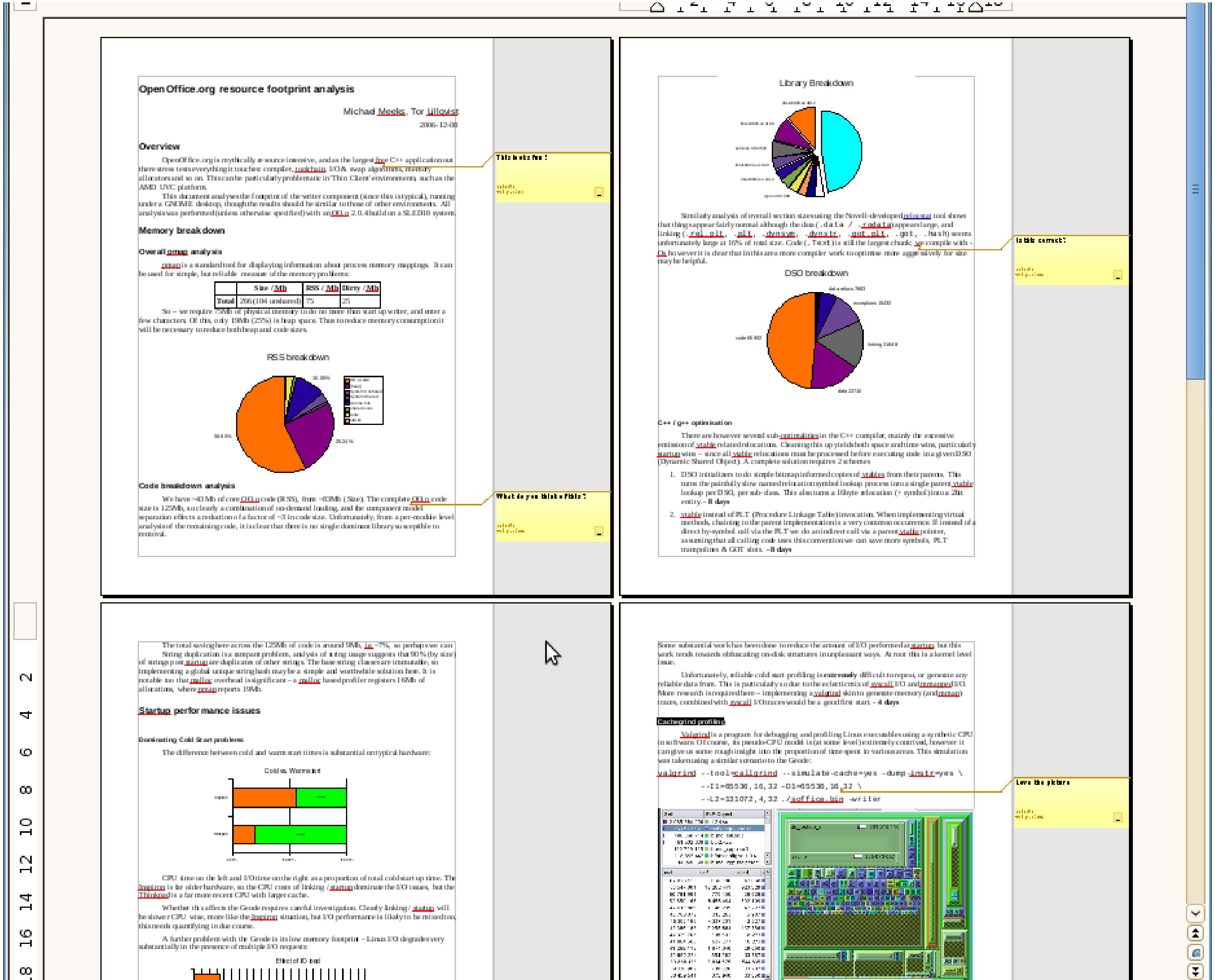
INWARD REPORT

(SHIPS AND BOATS OTHER THAN SMALL CRAFT)

PART A: ARRIVAL DETAILS		
Name of ship		Port of arrival
Date of arrival	Time of arrival	Port arrived from
Name of person in charge		Name and address of ship's agent in New Zealand
IMO number		
Gross register tonnes		
Period of stay in New Zealand		
IF CARRYING CARGO, BRIEF DESCRIPTION		

LIST SUBSEQUENT PORTS OF CALL WITHIN NEW ZEALAND (Please record on a separate sheet and attach if additional space is required.)	
Next New Zealand ports	Estimated date and time of arrival
1.	
2.	
3.	
4.	

Multi-page view & notes



OpenOffice.org resource footprint analysis

Michael Meeks, Tor Lillqvist
2006-12-08

Overview

OpenOffice.org is mythically resource intensive, and as the largest free C++ application out there stress tests everything it touches: compiler, toolchain, I/O & swap algorithms, memory allocators and so on. This can be particularly problematic in 'Thin Client' environments, such as the AMD UVC platform.

This document analyses the footprint of the writer component (since this is typical), running under a GNOME desktop, though the results should be similar to those of other environments. All analysis was performed (unless otherwise specified) with an OO.o 2.0.4 build on a SLED10 system.

This looks fun !

Memory breakdown

Overall pmap analysis

pmap is a standard tool for displaying information about process memory mappings. It can be used for simple, but reliable measure of the memory problems:

	Size / Mb	RSS / Mb	Dirty / Mb
Total	266 (104 unshared)	75	25

So – we require 75Mb of physical memory to do no more than start up writer, and enter a few characters. Of this, only 19Mb (25%) is heap space. Thus to reduce memory consumption it will be necessary to reduce both heap and code sizes.

RSS breakdown

Code breakdown analysis

We have ~40Mb of core OO.o code (RSS), from ~100Mb (Size). The complete OO.o code size is 125Mb, so clearly a combination of on-demand loading, and the component model separation effects a reduction of a factor of ~3 in code size. Unfortunately, from a per-module level analysis of the remaining code, it is clear that there is no single dominant library susceptible to removal.

What do you think of this ?

Library Breakdown

Similarly analysis of overall section sizes using the Novell-developed relocstat tool shows that things appear fairly normal although the data (.data / .rodata) appears large, and linking (.rel.plt, .plt, .dynsym, .dynstr, .got.plt, .got, .hash) seems unfortunately large at 16% of total size. Code (.Text) is still the largest chunk; we compile with -Os however it is clear that in this area more compiler work to optimise more aggressively for size may be helpful.

Is this correct?

DSO breakdown

C++ / g++ optimisation

There are however several sub-optimalities in the C++ compiler, mainly the excessive emission of vtable related relocations. Cleaning this up yields both space and time wins, particularly startup wins – since all vtable relocations must be processed before executing code in a given DSO (Dynamic Shared Object). A complete solution requires 2 schemes:

1. DSO initializers to do simple bitmap informed copies of vtables from their parents. This turns the painfully slow named relocation symbol lookup process into a single parent vtable lookup per DSO, per sub-class. This also turns a 16byte relocation (+ symbol) into a 2bit entity. – **8 days**
2. vtable instead of PLT (Procedure Linkage Table) invocation. When implementing virtual methods, chaining to the parent implementation is a very common occurrence. If instead of a direct by-symbol call via the PLT we do an indirect call via a parent vtable pointer, assuming that all calling code uses this convention we can save more symbols, PLT trampolines & GOT slots. – **8 days**

The total saving here across the 125Mb of code is around 9Mb, ie. ~7%, so perhaps we can

String duplication is a rampant problem, analysis of string usage suggests that 50% (by size) of strings post startup are duplicates of other strings. The base string classes are immutable, so implementing a global unique string hash may be a simple and worthwhile solution here. It is notable too that malloc overhead is significant – a malloc based profiler registers 16Mb of allocations, where pmap reports 19Mb.

Startup performance issues

Dominating Cold Start problems

The difference between cold and warm start times is substantial on typical hardware:

Cold vs. Warm start

CPU time on the left and I/O time on the right as a proportion of total cold start up time. The Inspiron is far older hardware, so the CPU costs of linking / startup dominate the I/O issues, but the Thinkpad is a far more recent CPU with larger cache.

Whether this affects the Geode requires careful investigation. Clearly linking / startup will be slower CPU wise, more like the Inspiron situation, but I/O performance is likely to be mixed too, this needs quantifying in due course.

A further problem with the Geode is its low memory footprint – Linux I/O degrades very substantially in the presence of multiple I/O requests:

Effect of IO load

Some substantial work has been done to reduce the amount of I/O performed at startup, but this work tends towards obfuscating on-disk structures in unpleasant ways. At root this is a kernel level issue.

Unfortunately, reliable cold start profiling is **extremely** difficult to repeat, or generate any reliable data from. This is particularly so due to the eclectic mix of syscall I/O and mmapped I/O. More research is required here – implementing a valgrind skin to generate memory (and mmap) traces, combined with syscall I/O traces would be a good first start. - **4 days**

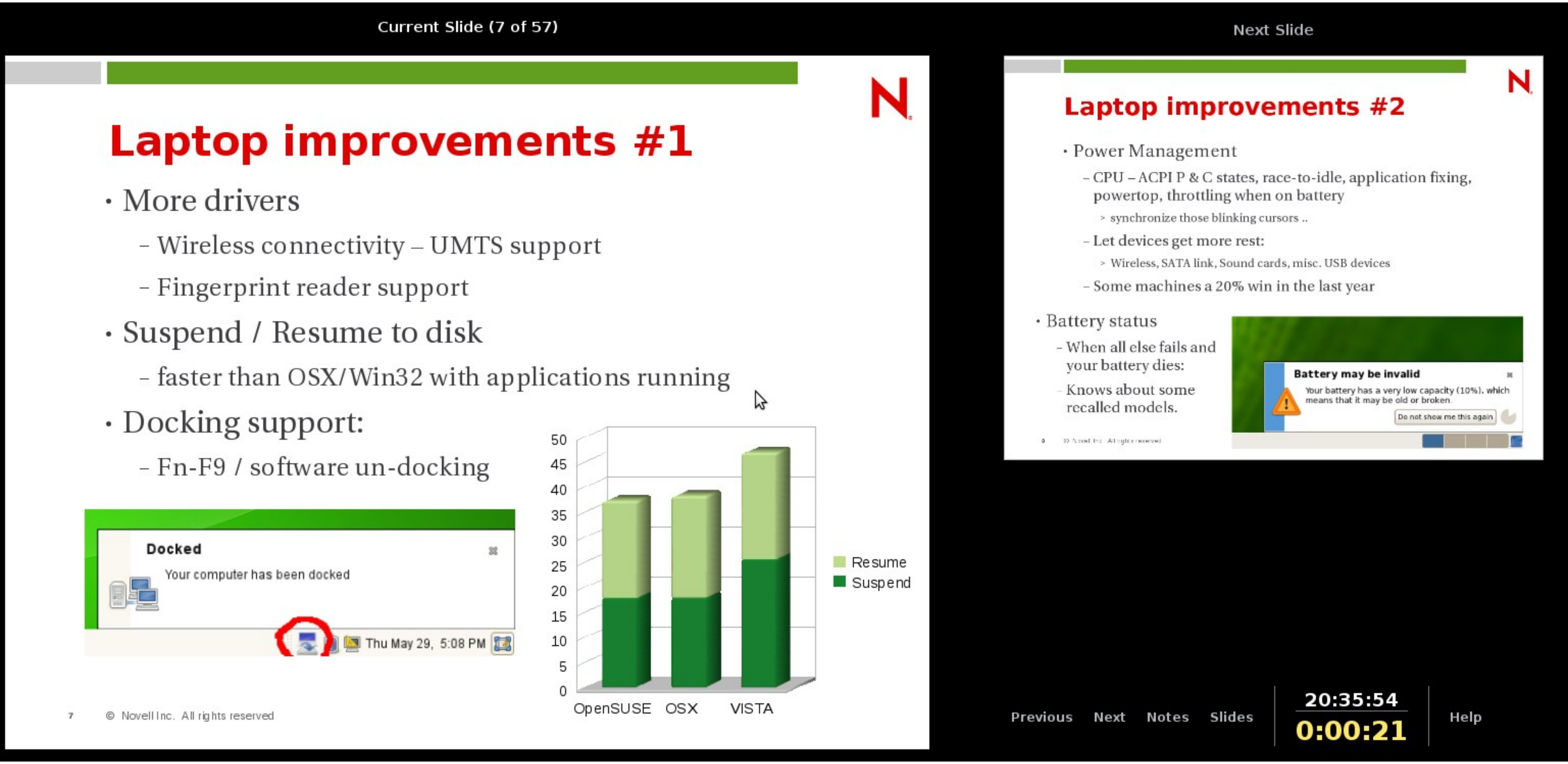
Cachegrind profiling

Valgrind is a program for debugging and profiling Linux executables using a synthetic CPU in software. Of course, its pseudo-CPU model is (at some level) extremely contrived, however it can give us some rough insight into the proportion of time spent in various areas. This simulation was taken using a similar scenario to the Geode:

```
valgrind --tool=callgrind --simulate-cache=yes -dump-instr=yes \
         --I1=65536,16,32 -D1=65536,16,32 \
         --L2=131072,4,32 ./soffice.bin -writer
```

Love the picture

Infamous Presenter view ...



Current Slide (7 of 57)

Laptop improvements #1

- More drivers
  - Wireless connectivity – UMTS support
  - Fingerprint reader support
- Suspend / Resume to disk
  - faster than OSX/Win32 with applications running
- Docking support:
  - Fn-F9 / software un-docking

Docked

Your computer has been docked

Thu May 29, 5:08 PM

7 © Novell Inc. All rights reserved

Next Slide

Laptop improvements #2

- Power Management
  - CPU – ACPI P & C states, race-to-idle, application fixing, powertop, throttling when on battery
    - synchronize those blinking cursors ..
  - Let devices get more rest:
    - Wireless, SATA link, Sound cards, misc. USB devices
  - Some machines a 20% win in the last year
- Battery status
  - When all else fails and your battery dies:
  - Knows about some recalled models.

Battery may be invalid

Your battery has a very low capacity (10%), which means that it may be old or broken.

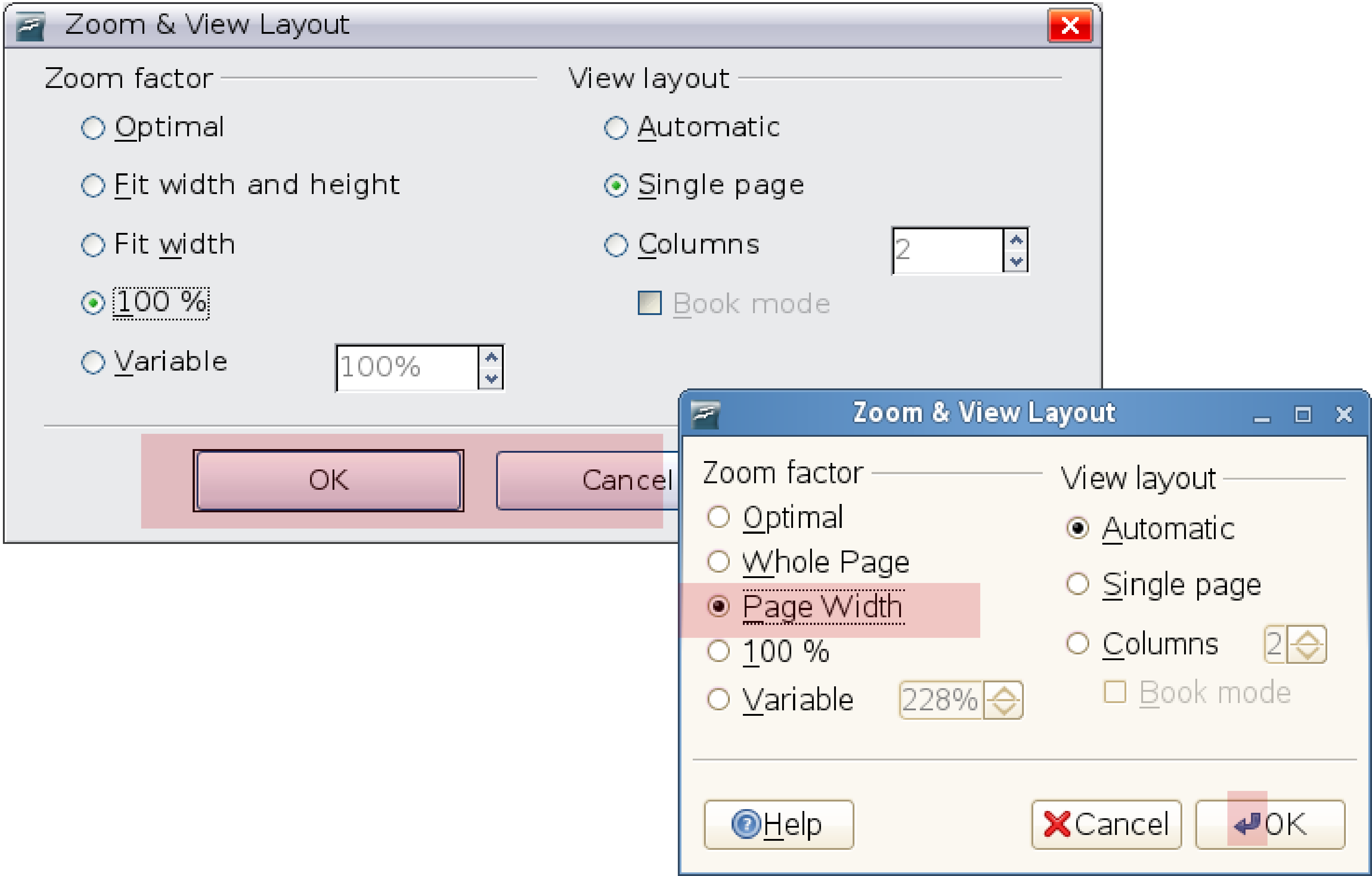
Do not show me this again

8 © Novell Inc. All rights reserved

20:35:54

0:00:21

UI Layout ... (demo)



Zoom & View Layout

Zoom factor

- ○ Optimal
- ○ Fit width and height
- ○ Fit width
- ● 100 %
- ○ Variable 100%

View layout

- ○ Automatic
- ● Single page
- ○ Columns 2
- □ Book mode

Zoom & View Layout

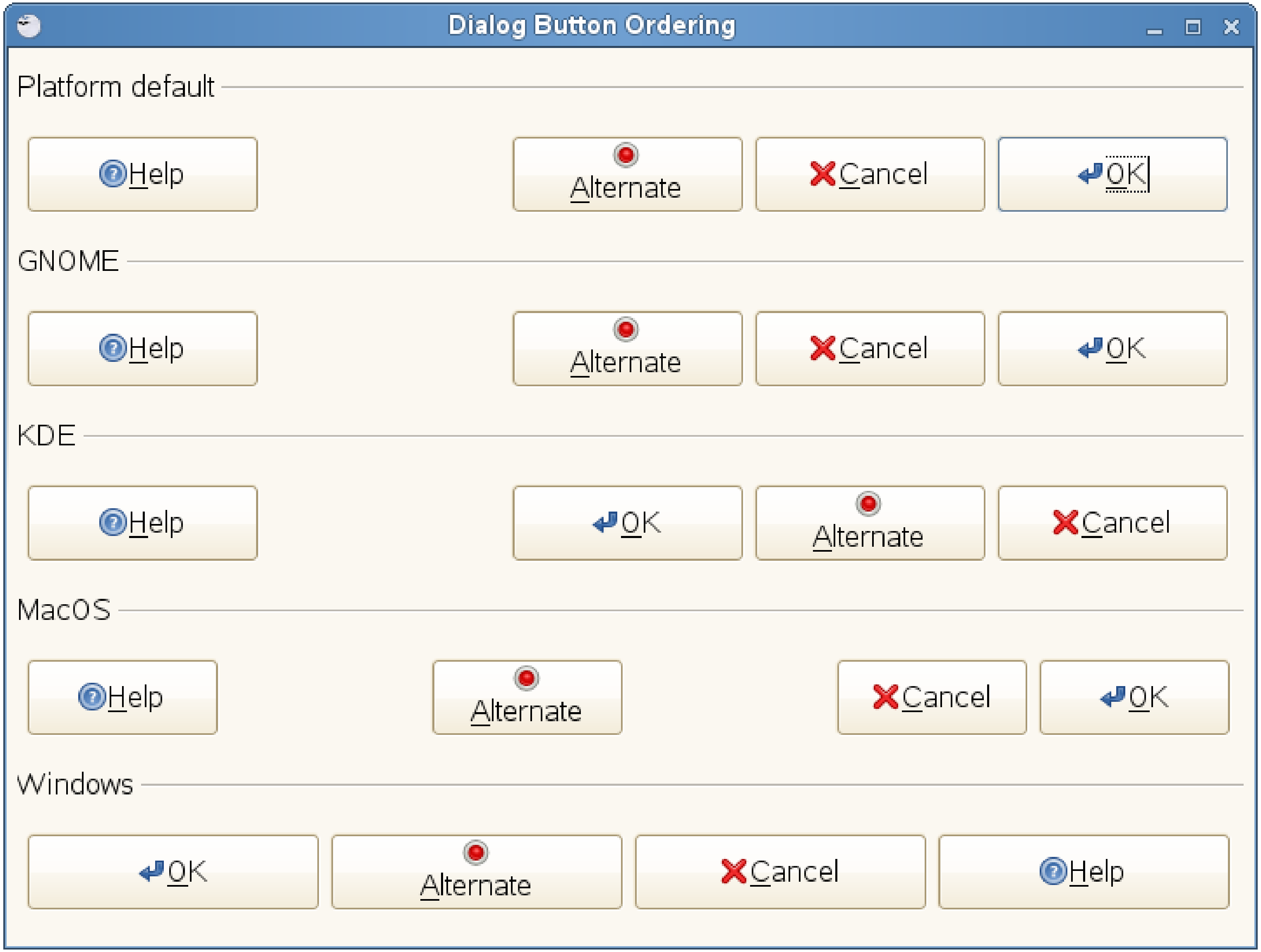
Zoom factor

- ○ Optimal
- ○ Whole Page
- ● Page Width
- ○ 100 %
- ○ Variable 228%

View layout

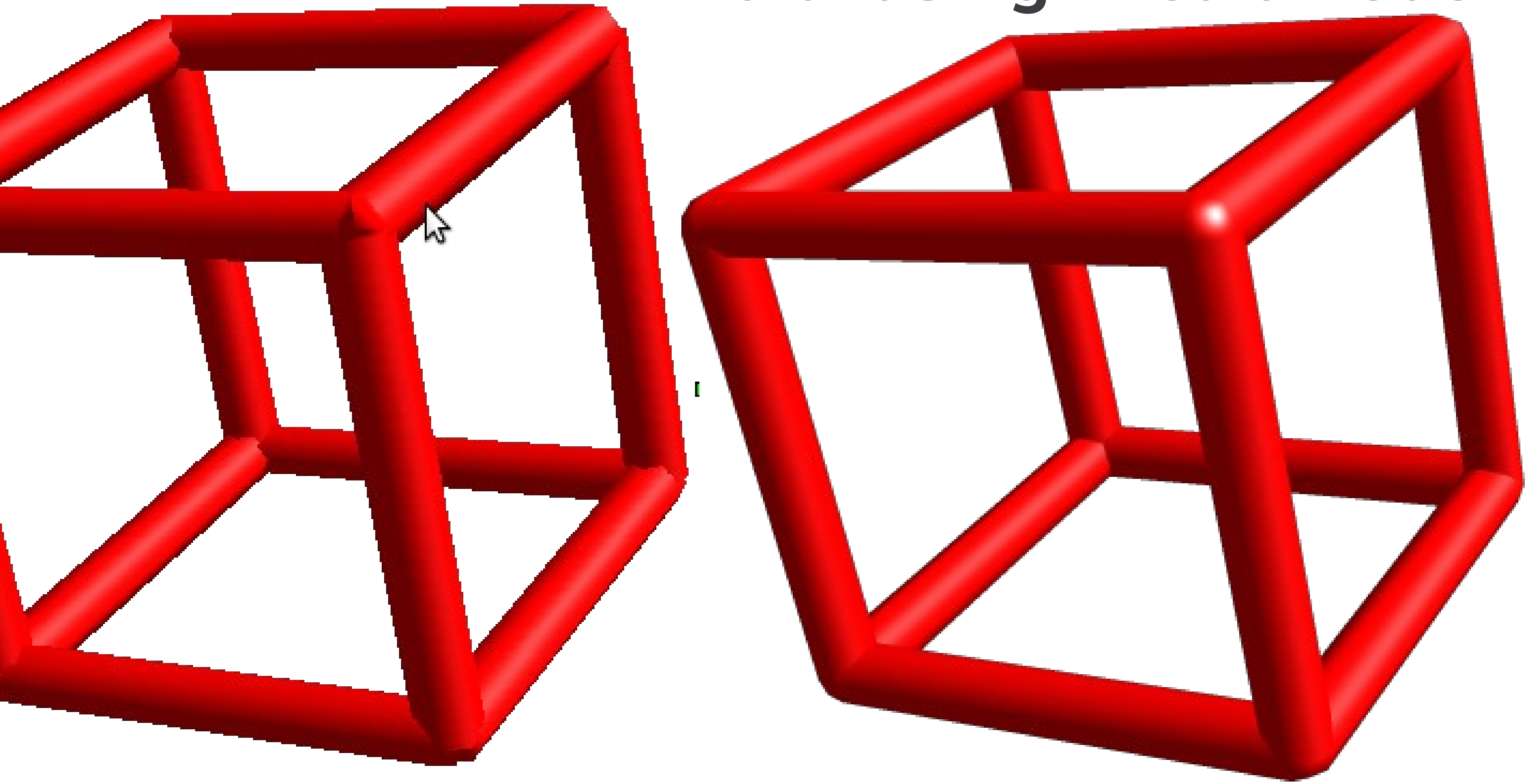
- ● Automatic
- ○ Single page
- ○ Columns 2
- □ Book mode

Platform button ordering ...



Drawing layer wins ...

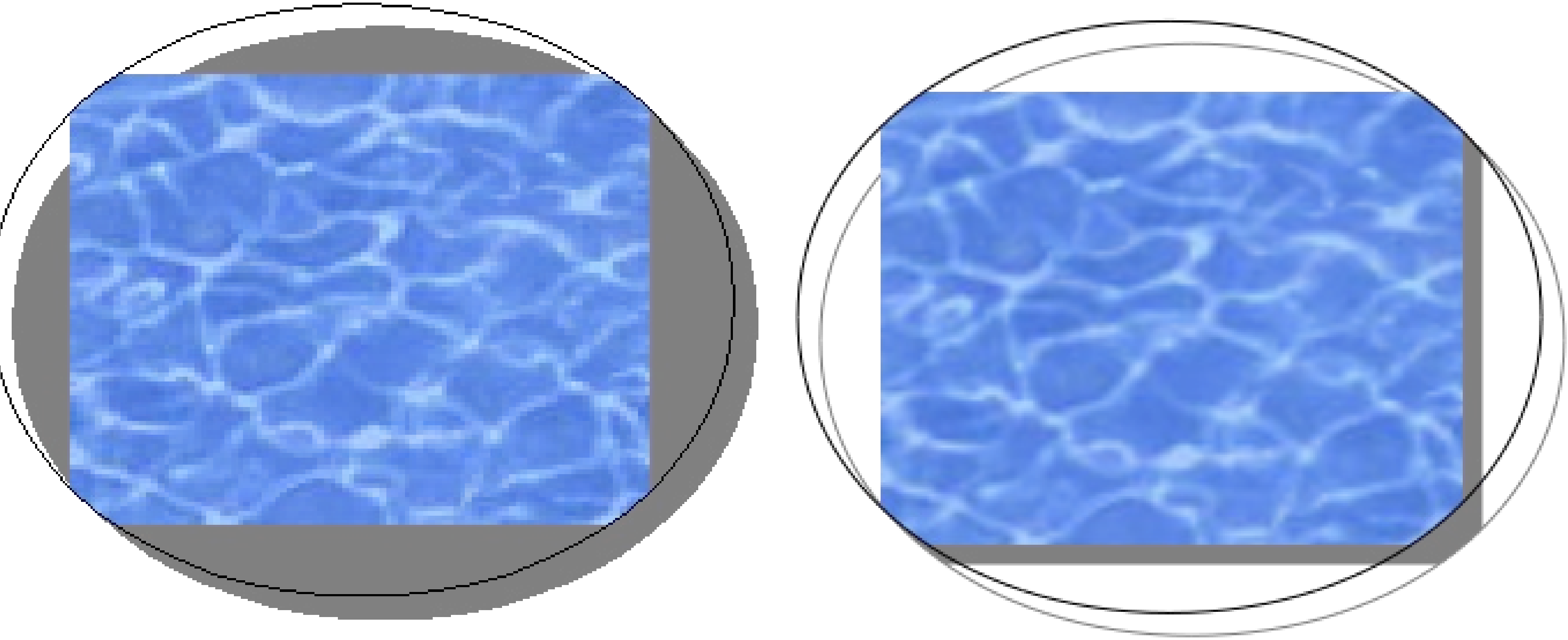
Anti-aliasing in edit mode



Correct caps / joins

More drawing layer wins ...

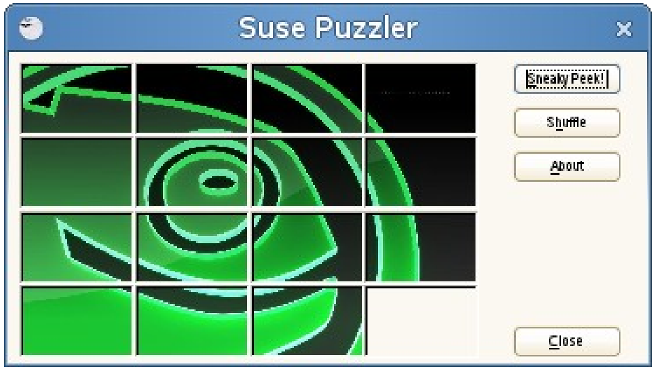
Real circles ...



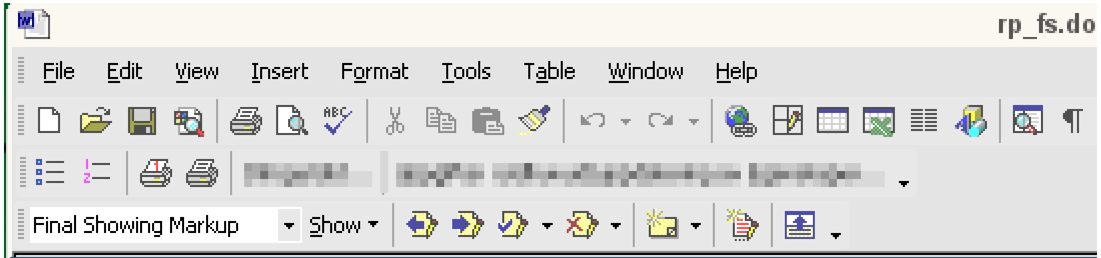
Real shadows ...

VBA, a language to love … (?)

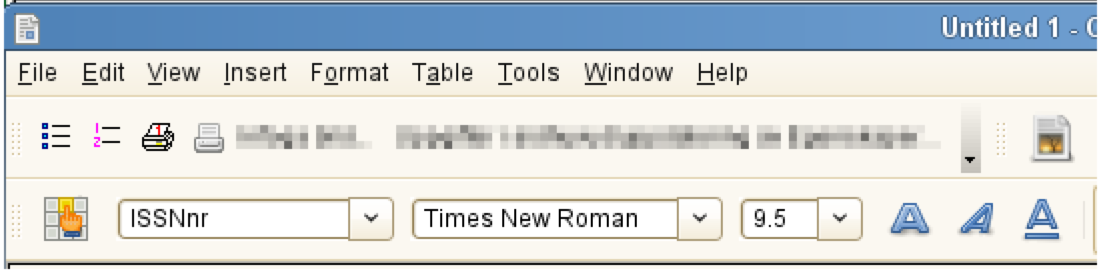
- New features coming: **(Demos)**
  - **Word macro support**
  - Improved User forms support
  - Better Calc support
  - Custom Menu / Toolbar support:



MS Word:



Writer:



VBA: the competition ...

- Microsoft Office: dropping VBA support on OS/X
  - Office12 OS/X – 9 months later...
  - http://www.schwieb.com/blog/2006/08/08/saying-goodbye-to-visual-b

The removal of VB means that **existing macros** in Office documents will be round-tripped across file open and save, but you will not be able to edit them and you will **not be able to run them on the Mac**. ... The MacBU is **very aware** of the **pain this decision will cause** for users, consultants, and **enterprise organizations**. I've personally seen the phrases **"apoplectic with rage"** and **"absolutely livid"** in two emails that crossed my inbox. ... As I mentioned in an earlier post, **GCC** is very picky about code meeting the current standards and the VBE code most certainly does not. ... VBA, on the other hand, is incredibly **difficult to port to Intel** ... Lastly, we have Forms. **Forms is** also C++, but is backed by **several thousand lines of gnarly custom assembly.** This assembly 'allows' the C++ code to swap object virtual function tables and individual member function pointers between objects on the fly, to essentially do very fast object morphing. ...

StarBasic vs. VBA ...

- Guess which – an interactive macro game:
  - [Examples from VbaStarBasicXref.sxi]

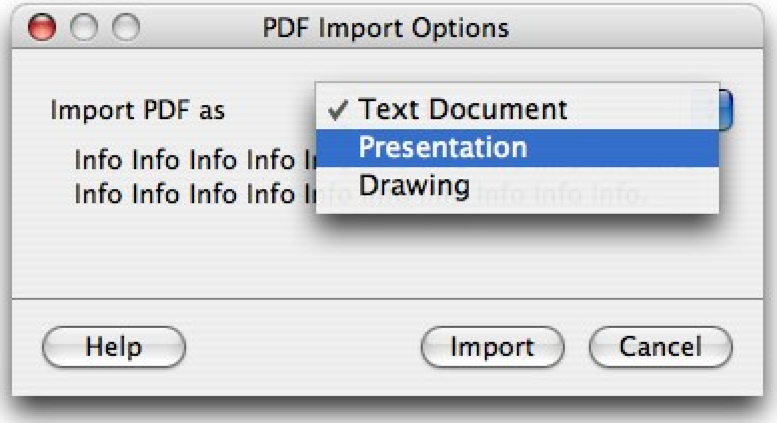
```
1. wksh = ThisComponent.CurrentController.ActiveSheet
vs.
   wksh = ActiveWorksheet


2. msgbox WorksheetFunctions.Average(Range("A1:A5"))
vs.
   Dim oSheet, FuncService
   FuncService = createunoservice("com.sun.star.sheet.FunctionAccess")
   oSheet = ThisComponent.CurrentController.ActiveSheet
   msgbox FuncService.callFunction(“AVERAGE”, _
        array(oSheet.getCellRangeByName("A1:A5")))
```

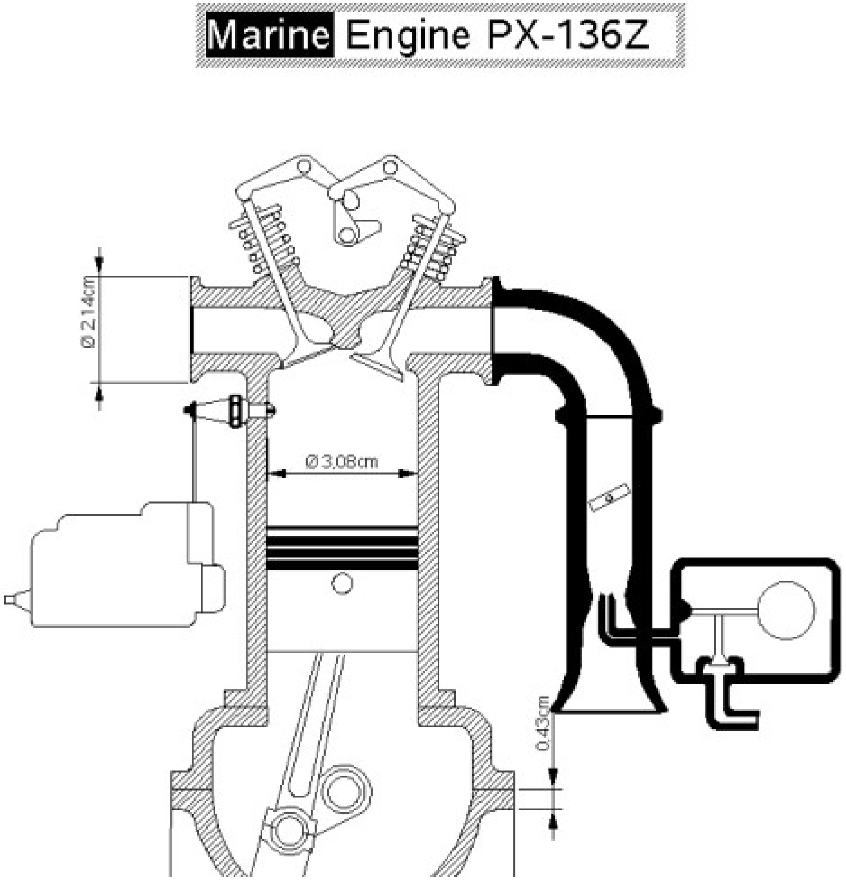
- Lots more examples ... - but no 'Replace' issue.

PDF Import ...

- Also sadly an extension ...



Import PDF as ✓ Text Document



PDF-Import - Draw

Under-cover curios …

UNO – bliss & agony ...

- UNO: an excellent / the best component model
- Unfortunately: over-reaches in scope ...
  - Automatic bindings for scripting languages
  - Interfaces for core application development
  - Hyper-granular thread safety mechanism
  - Eternal ABI / behavioral stability point ...
- The result: complication ...
  - baroque & unusable interfaces for scripting
    - not granular & friendly enough
  - horrific implementation problems wrt. Thread-safety
    - Interfaces much too granular
- Hard-to-use thing – and worse a **Cult** to boot.

OpenDocument / Office Open XML

- Structure: Non-mixed   Or mixed ...

```
MS OpenXML
<w:r>
  <w:rPr>
    <w:b />
  </w:rPr>
  <w:t>this is bold</w:t>
</w:r>
```

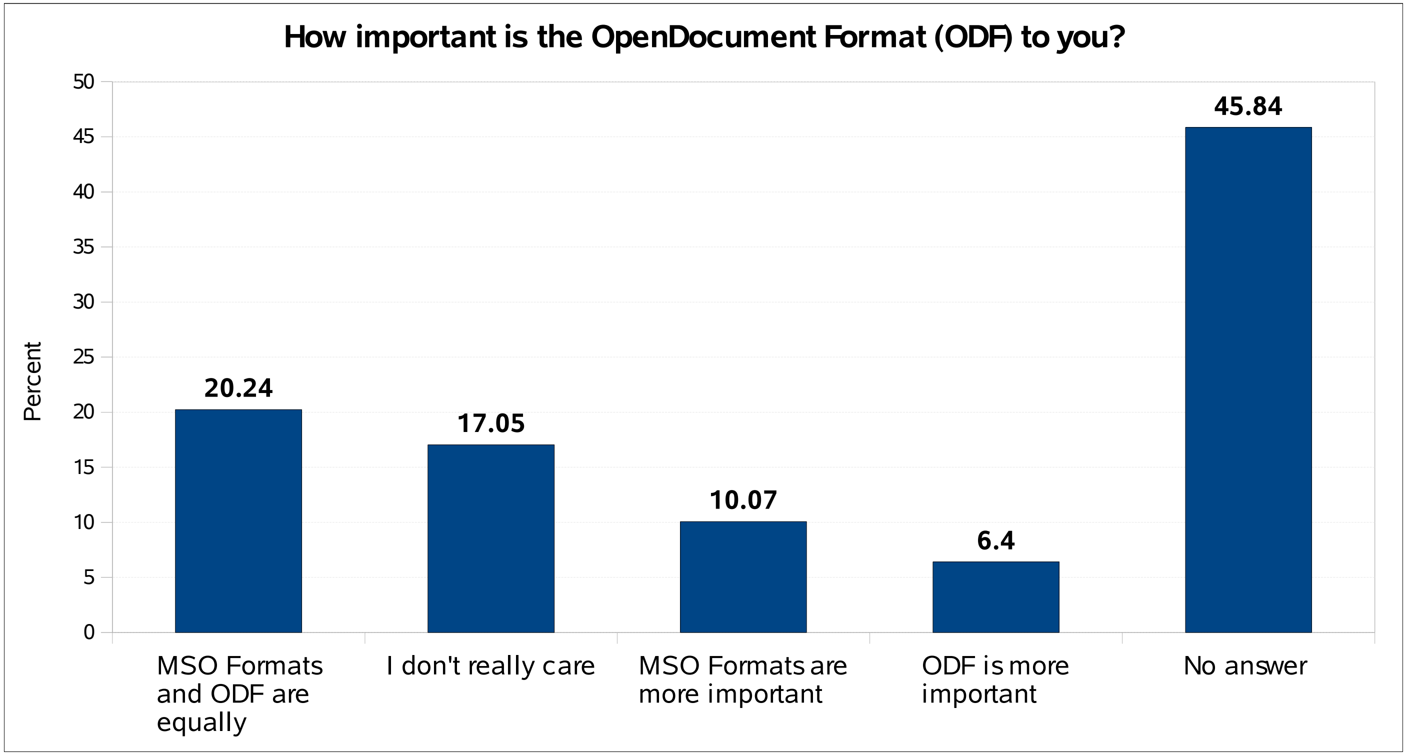
```
OpenDocument
<text:span text:style-
name="Strong_20_Emphasis">
    this is bold
</text:span>

XHTML
<b>this is bold</b>
```

- The good thing about standards is ...
- OpenOffice – over OpenStandards: Liberté
- Free software development should, provides an open, permanently recorded development process, open to all participants
  - You can also use the software now & in the future ...

What about the users ?

- OO.o User Survey: 160k users to end of Dec 2008

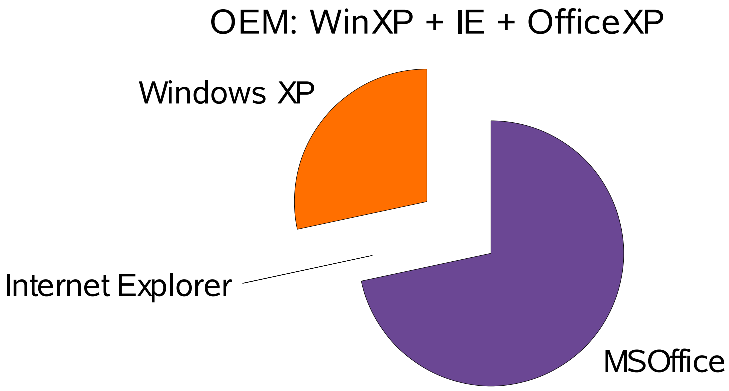


- Education wrt. Free Software required ...

Getting involved …

Why OO.o is so important

- The price gradient:
  - 'Normal' people using it today
  - runs on Win32
  - N x millions of users
  - big deployments...



$12bn in 2005
30% of revenue

- Developer bang for buck
  - where is Freedom best served ?
  - what Free software do people see 1st ?
- Personal impact
  - you can easily become 10+% of the calc team
  - and make a huge difference
  - The code is not **that** bad: you can do it ...

Some sample tasks ...

- Cross compilation – from Linux to …
  - Win32 – MingW – **lots** of simple, scattered fixing
  - OS/X – we can use that.
  - Cmake migration from dmake
- Build-bot setup / maintenance
- Layout work – polishing & porting dialogs
- VBA – more object module bits – just 1 method …
- Calc scalability – real men's problems …
- QA, bug triage, artwork, translation etc ...

Conclusion / Q&A

- OpenOffice is a mixed bag:
  - improving in some areas, from a low base.
  - static / regressing in others
- It needs your help & love to make it the perfect suite.
- Thanks – to all the hackers that stay the course, and still contribute …
- Mail me, to get involved.

Oh, that my words were recorded, that they were written on a scroll, that they were inscribed with an iron tool on lead, or engraved in rock for ever! I know that my Redeemer lives, and that in the end he will stand upon the earth. And though this body has been destroyed yet in my flesh I will see God, I myself will see him, with my own eyes - I and not another. How my heart yearns within me. - Job 19: 23-27