

The wonderful world of:



LibreOffice

The Document Foundation

Or – How we will fix your office suite

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 & disclaimer ...

- I am not speaking for the Document Foundation
 - These views are mine, but you're welcome to share them.
- The What and Why of LibreOffice
 - How is it going ?
 - What are we doing ?
 - What caused the change ?
- Ownership aggregation
 - Emerging problems & Legal foobaa's
- Plumbing / tooling fun
- Application / Conclusions

What / Why of LibreOffice

What is LibreOffice ?

- The wave of the future
 - a beautiful, Office suite we can be proud of (in due course) backed by a real, open community
- But the Web will swallow us all !!?!?
 - full function off-line == re-write all in Javascript
 - Best of luck with this project: 8 million lines later
 - eg. Conference WiFi & 3G is extremely cluttered
 - I would like to show you these slides today
 - but they are on a server I can't get to
 - Not: Rent vs. Buy cloudness ? – *Buy Furcate, Rent Collectors ?*
- My Phone is an 80's super-computer anyway ...
 - *seen my vt220 app ?*

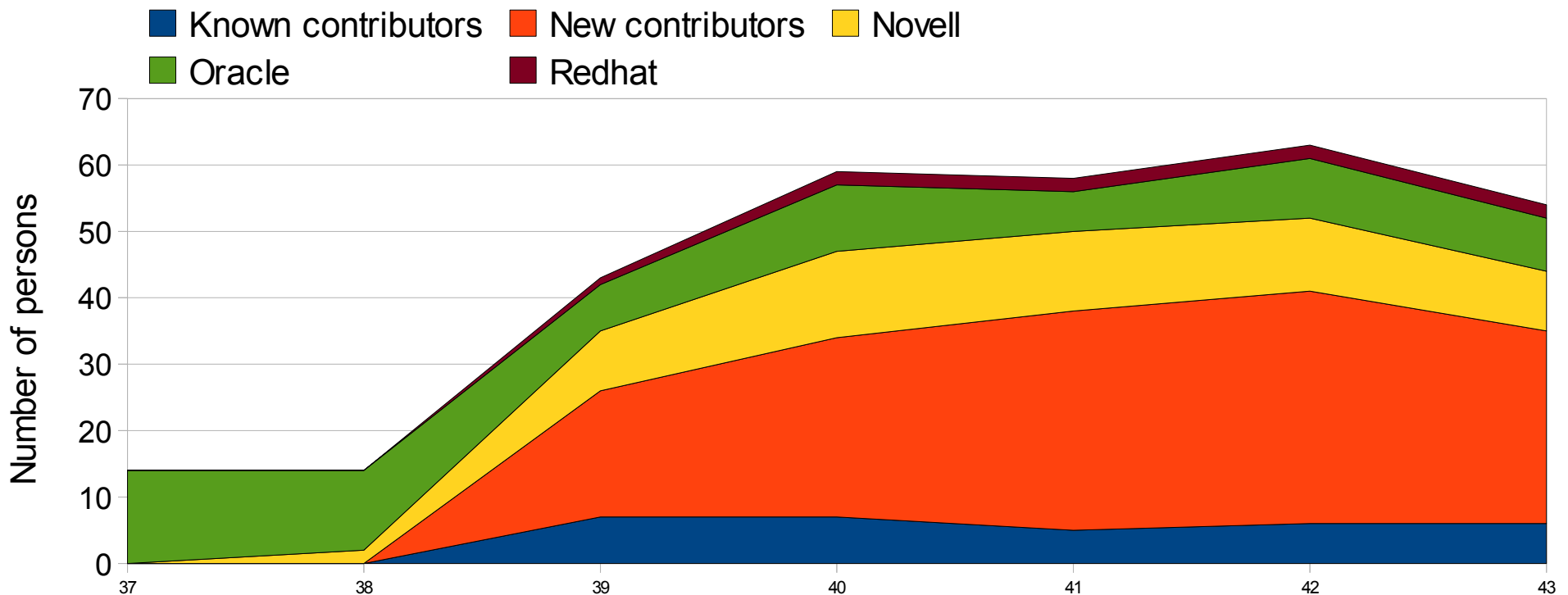
Why / Where LibreOffice ?

- ▶ Ten years after the promise of a foundation ...
 - ▶ “a foundation is a great idea ... the time is not yet ripe ... perhaps in another three years” etc. etc.
 - ▶ Patience – a virtue, but not an inexhaustible resource.
- ▶ Some truly fantastic 'opportunities' for improvement
 - ▶ Vendor neutrality / no copyright assignment barrier
- ▶ The need to (finally) “do it right”
 - ▶ a real Free Software / hackers project
 - ▶ Volunteers + RedHat + Novell + Debian ...
 - ▶ FSF + OSI support, boycott + Novell support etc.
 - ▶ Freedesktop hosting
 - ▶ LGPLv3+ / MPL for new code ...
- ▶ An idea – whose time has finally come ...

How is it going ...

Extremely well / as of now:

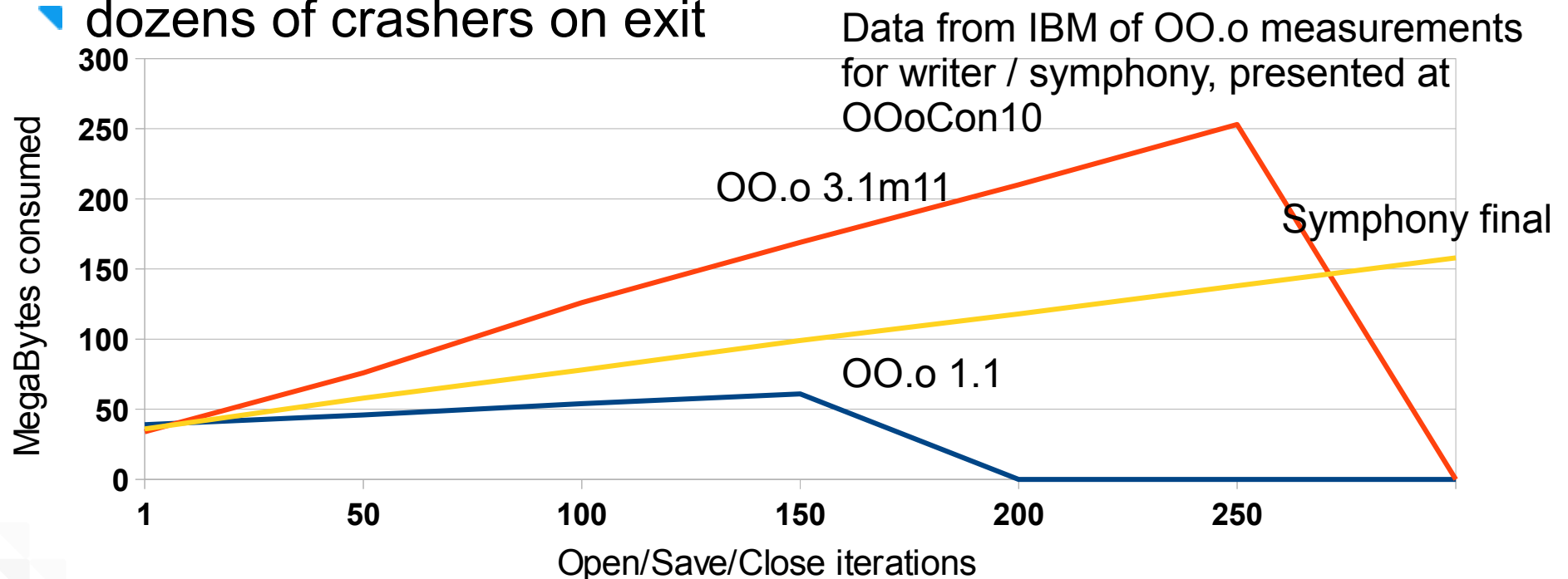
- 50 entirely **new** code contributors with included patches, **27** translators
- another 10+ arriving each week
- 100's of KLOC of real patch [3+million line diff to branch-point] (meaningless?)
- But – Oracle won't join us in name (yet) ... though they contribute indirectly.



What are we doing ?

- Pay down the vast technical debt we inherited
 - an ancient and gnarled, 20+yr old beast
 - organic growth, without refactoring, a lack of code review etc.
- eg. the opt-in approach to lifecycle management:

- dozens of crashers on exit



What are we doing ? #2

- ▼ Translating the source code:

- ▼ 100k+ LOC of German comments, after 10 years of 'open'

- ▼ *Some virtually useless: (CUPS style)*

```
//      Beschreibung:      textuelle Beschreibung ermittel  
String SwDocStyleSheet::GetDescription(SfxMapUnit eUnit)
```

- ▼ *Others somewhat more useful:*

```
|*      Wenn bereits ein passendes Format existiert, so wird dies  
|*      zurueckgeliefert.  
-|* Ersterstellung      MA 22. Sep. 92  
-|* Letzte Aenderung JP 08.05.98  
+|* Created:           MA 22. Sep. 92  
+|* Last modified:     JP 08.05.98
```

- ▼ *And more that are useful (we think)*

- ▼ Fantastic “German comment detector” from Miklos

- ▼ Re-uses existing LO code for language detection :-)

What are we doing ? #3 – Revision Control ...

- ▼ OpenOffice.org had a chequered history RCS wise
 - ▼ collab.net's CVS – multiple-hours to tag the repository ...
 - ▼ After years of pain – rejected *git* for temporary move to svn
 - ▼ Then moved to mercurial – history routinely lost
- ▼ Thus – a reliable workaround: the in-code RCS ...

- ▼ #if 0 – is your friend ! (good for 10's of kloc)
- ▼ If in doubt, just comment it out (huge chunks of code):

```
//UNUSED2008-05          if (nCount==1)
//UNUSED2008-05          return 0;    // leer
```

- ▼ Annotate sections you change carefully: etc. etc.

```
// --> OD 2005-12-12 #i27301# - use new type definition for parameter...
```

- ▼ LibreOffice – uses git [!]
 - ▼ We bin insubstantial or un-necessary comments to improve readability.

What are we doing ? #4

- Making it easy to contribute
 - Google search: *libreoffice easy hacks*
 - Giving permission for change: *we value your changes!*
- Creating unit tests (run during build) – Caolan
 - Currently next-to-no worthwhile unit tests
- Removing VOS (deprecated a decade ago) – **Norbert**
- Merging outstanding Debian, RedHat, Novell patch-sets
- Misc. community features eg. GSOC bits
 - RTF export, reducing code duplication, bug fixing – Miklos
 - In-line, interactive formula editing – Jonas
- Bug, warning, lint error, cppcheck fixing ...

An example ...

▸ John Lee Castle

- arrives on list to help debug a (star)basic crash
- 2 days later – first set of 'warning' patches merged
- Sadly beaten to the word-count feature, but ...
- digs in bugzilla to find & 1 week in fixes Issue 76852:

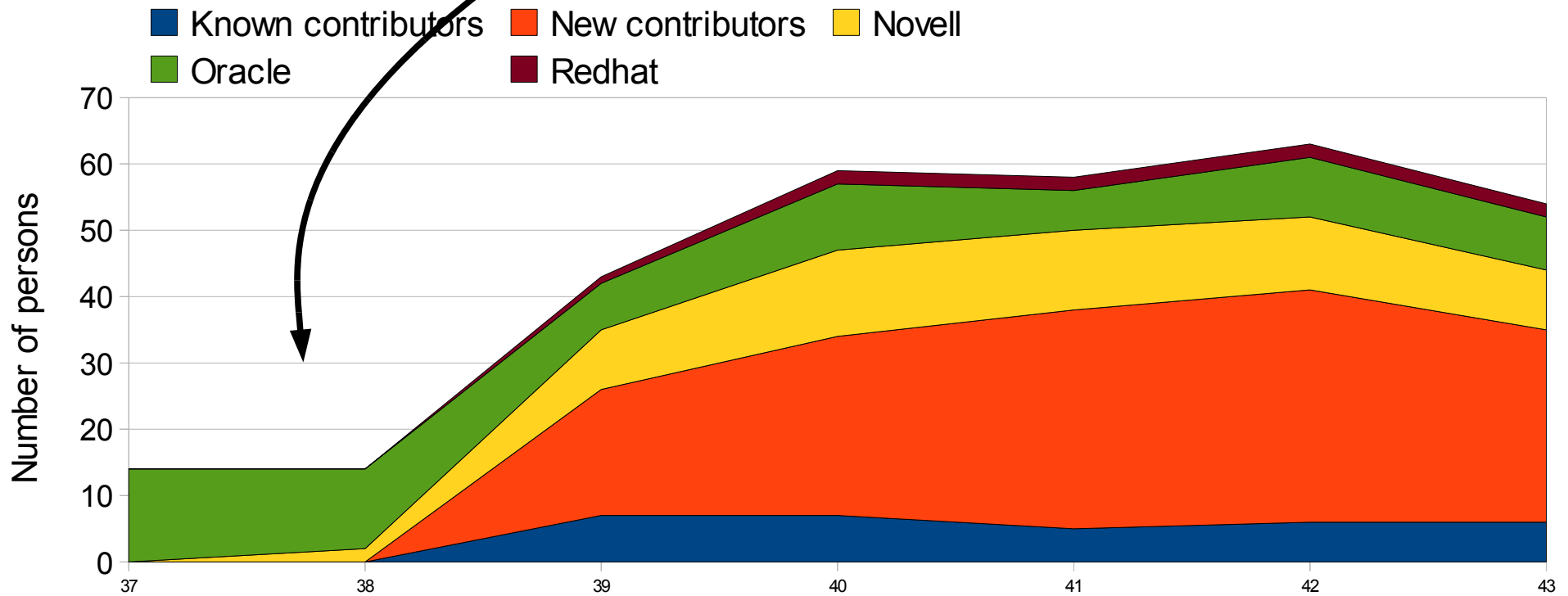
```
Opened: Wed May 2 12:28:00 +0000 2007
```

```
Run this Basic macro:
```

```
Dim s as single  
s = 999999  
print s ' display : 999999  
s = s+1  
print s ' display : 1 *****  
incorrect ***  
s = s+1  
print s ' display : 1.000001E6
```

What changed here ?

*Day #1 nearly same code
Is it just an 'easy hacks' page ?*



This slide deliberately left blank

Vendor neutrality / no assignment of ownership ...

- ▼ An emerging subtle threat to software freedom
 - ▼ Assigning ownership to companies
 - ▼ Not always 'copyright' – sometimes just a public domain contribution, leaving an exclusive corporate owner of the complete work.
 - ▼ Often assignment with a fall-back license: perpetual, irrevocable, worldwide, no-charge, royalty-free, unrestricted ...
- ▼ Why should I care ?
 - ▼ Traditionally people mostly cared about (Free) licenses (good)
- ▼ I will argue that you need to care about others' terms too
 - ▼ Assignment creates a barrier of delay and conflict
 - ▼ It disrupts a delicate hacker / suit-wearer balance
 - ▼ It entrenches corporate control, and excludes competitors
 - ▼ It creates FUD around free-software licenses & practices

Assignment creates a barrier of delay, and conflict

- ▶ Even in the (much more defensible) case of assignment to a well governed non-profit – there are (*historical*) problems:

Why are there two Emacsen? [xemacs and emacs]

Hysterical reasons. See the public flames if you must know. Currently the largest problems is that **Sun owns the copyright** for large part of the code. It is copyleft, but **they refuse to sign** over the changes to the FSF. Some people think it is good to have an Emacs variant where you **can contribute code without signing any papers**. Feel free to disagree.

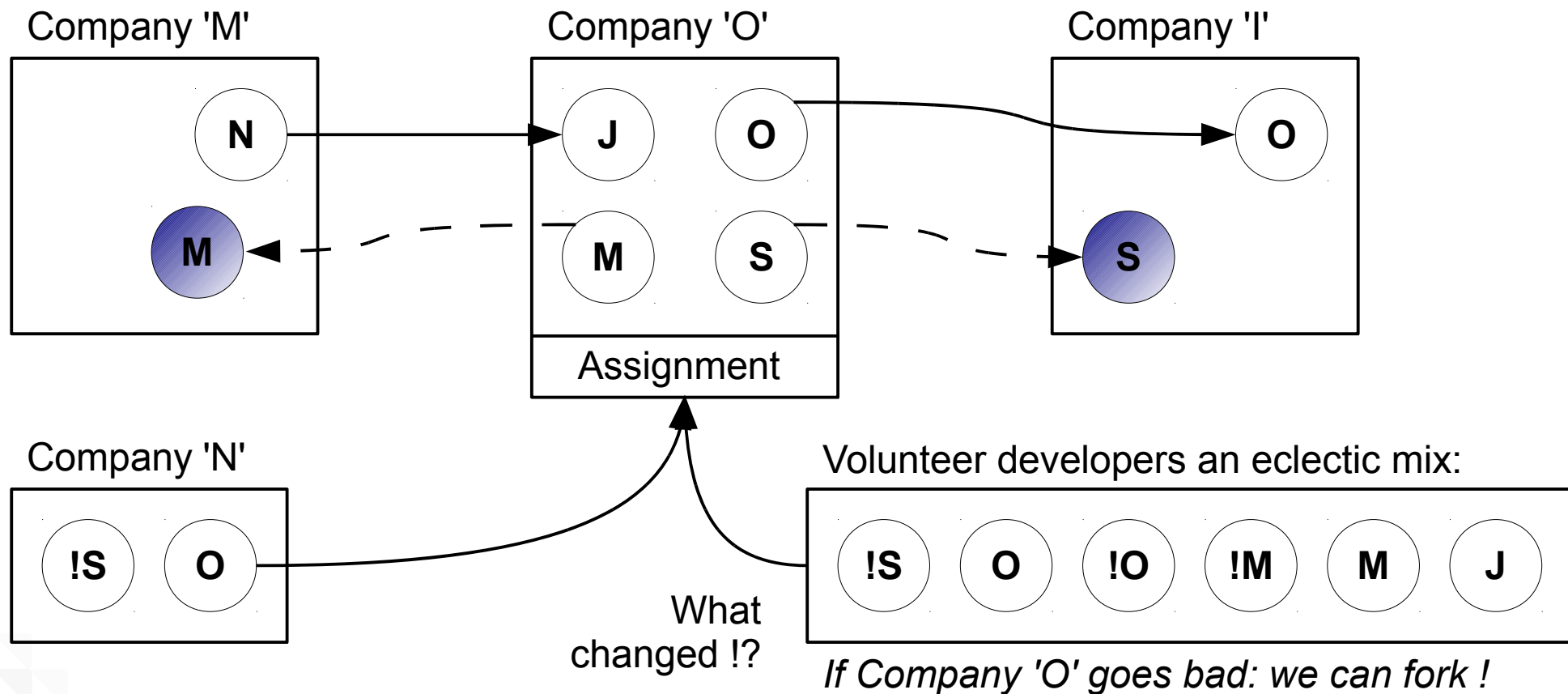
- ▶ Ergo, some people, sometimes think this is a bad thing.
- ▶ **The Best of Strangers**
 - ▶ A brilliant research paper on privacy expectations
 - ▶ Strong suspicion that handing people a dense and comprehensive legal paper to sign on day-zero can badly hurt trust.
- ▶ **Systemd**: “*We value your contributions and hence do not require copyright assignment*” - we welcome you as a peer ... immediately

Disrupting the Hacker / Suit balance ...

- ▶ The very existence of this debate...
 - ▶ Your pet project: is already owned by pointy haired boss
 - ▶ He doesn't listen to advice from Dilbert in linear time
- ▶ Technical and relational understanding diminish upwards
 - ▶ All that good work you did means nothing vs. hard \$
- ▶ Importing the P&L / competitive angst into our community
 - ▶ Many people I call friends at RedHat, no doubt our sales and business people compete aggressively
- ▶ Technology / stack decision making
 - ▶ good results from considering ownership / corporate alliances at a business level first ? *This is our swamp ! We build here !*
 - ▶ most inter-corporate agreements are confidential: death of community transparency ?

Disrupting the Hacker / Suit balance ...

- ▶ Enleveraging synergistic technological progress:
 - ▶ In a proprietary and confidential fashion:



Entrenching corporate control: Excluding others ...

- ▼ **Disclaimer: this is a project I know – nothing – about.**
- ▼ The 'apparently' open project
 - ▼ **OpenSPARC** – ~GPLv2 licensed + 'Shared' CA
 - ▼ *All contributors to OpenSPARC are required to sign the Contributor Agreement ("CA") before contributing code to the community. This enables a single entity to **represent** the aggregated code base and gives the '**community**' flexibility to act as a whole to changing situations.*
- ▼ SPARC or OpenSPARC ?
 - ▼ Safest to proprietary license from Oracle ... ?
 - ▼ Needs a 'Welcome' sign with: “Academics, non-profits & partners welcome”

Entrenching corporate control: 'Open' Core

- ▶ Open Core (according to Lampitt)
 - ▶ core is GPL: if you embed the GPL in closed source, you pay a fee
 - ▶ [optional tech-support for a fee]
 - ▶ “*annual commercial subscription includes: **Indemnity**, ...*”
- ▶ Claims of **special** “*Indemnity*” are suspect:
 - ▶ Anyone can offer indemnity on Free Software
 - ▶ RedHat & Novell & many others do, for code they do not own
 - ▶ The '**special**' piece, has to be the ability to avoid the license everyone else has to accept...
 - ▶ *Enabled by ownership assignment*
- ▶ How do you market & sell an 'indemnity' product ?

FUD around Free-Software licensing & practise ...

- ▼ So why would people ever sign such a thing ?
- ▼ Some great arguments:
 - ▼ *Copyright assignment enabled us to move from the GPLv2 to the GPLv3*
 - ▼ *Better written as: we chose a non-plus license because we can.*
- ▼ Re-licensing:
 - ▼ Plus-licenses: GPL or later version, MPL, CDDL ...
 - ▼ Non-Plus: Apache (use CLA), EPL etc.
 - ▼ So liberal no-one cares: BSD, MIT, X11, etc.
- ▼ The 'standing' chestnut ...
- ▼ Deep seams of jurisdictional nonsense:
 - ▼ “Upper Silesian tax law makes Free-software contribution a felony ... ergo we need a custom Silesian CLA to be safe”

FUD around Free-Software licensing & practise ...

- ▼ But we **do** have a contributor agreement
 - ▼ The license we use:
- ▼ A beautiful symmetry:
 - ▼ **Inbound == Outbound** best in most cases
- ▼ I am not a Lawyer (sadly)
 - ▼ Richard Fontana (RedHat) who is, has a great talk on this:
 - ▼ <http://ref.fedorapeople.org/fontana-harmony.html>
 - ▼ Legal benefits of greater control sufficiently dubious to not justify harm done to the community by assignment
 - ▼ **Contributor Agreements seem to suggest that free software licenses are not good enough for the model.**
 - ▼ Ethical concerns: (unequal bargaining power)

More FUD: sustained investment ...

*“The only way companies can continue to invest in working on non-sexy projects is if they can **'own'** it”*

▼ **Or is it ?**

▼ Counter examples:

▼ almost our whole ecosystem is not owned by companies

▼ Linux, util-linux, udev, glibc, gcc, perl, python, Mozilla, Webkit, Apache, dbus, gtk+, GNOME, etc.

▼ *Personally ~all (cleanly written) software is quite sexy*

▼ **A counter-statement:**

▼ *“We must not let companies, who are uncertain of their continued investment, to exclude others from co-owning and building out key pieces of our stack”*

More FUD: the valuable MySQL model ...

“Sometimes the only way we can get Free software in a given space is to allow one company a monopoly of ownership and control there: We must not hurt the valuable MySQL model !

▼ **A counter-suggestion:**

*“We must not let companies wanting to dominate, own and exclude competitors from a part of our stack / ecosystem to **claim** they are doing something good for Free Software”*

- ▼ Consider where we could be with a true, organic alternative.
- ▼ The Sun acquisition dented the shine on this model too.

More FUD: no traction for non-CA projects

“If you look at projects with lots of traction, only assignment based projects are a success. We should be grateful to the companies that do all the heavy lifting (alone) there.”

▼ **A counter-suggestion:**

“Linux is the paradigmatic successful Free Software project. No corporate ownership aggregation. And there are many others. Indeed – a single company doing all the heavy lifting is a sign of risk, and failure is one acquisition away.”

Corporate Ownership trends ...

- ▼ (C) assignment projects get forked or re-written
 - ▼ xpdf → poppler, libart_lgpl → cairo, MySQL → MariaDB ...
- ▼ Evolution
 - ▼ Novell dropped (C) assignment -> LGPLv2/3
- ▼ Red Hat has dismantled several contribution agreements: SpaceWalk, FreeIPA, JBoss CLA under review, Fedora's is now optional
 - ▼ use in-bound equals out-bound in their place
- ▼ **GNOME** guidelines on assignment
 - ▼ demands outbound type requirements in CAs
 - ▼ patent licenses must apply to Free Software version
 - ▼ proprietary licenses must be non-discriminatory and be published

Misc. technical breather ...

Git – good but could be better ...

- ▼ Git clone – we have lots of bulky, irrelevant history

- ▼ Push from shallow clones ?

`--depth <depth>`

Create a shallow clone with a history truncated to the specified number of revisions. A shallow repository has a number of limitations (you cannot clone or fetch from it, **nor push from** nor into it), but is adequate if you are only interested in the recent history of a large project with a long history, and would want to send in fixes as patches.

- ▼ git-am foo.diff

“no valid header” ...

- ▼ Nothing applied – but transaction not auto-aborted
 - ▼ No warnings until 'git push': “git am in progress”

Valgrind – Dynamic Heap Analysis Tool (DHAT) ...

- Extremely fun new tool, analyses usage of each byte on heap: how long allocations were alive etc.

Aggregated access counts by offset:

```
[  0] 198918 198918 198918 198918 198918 198918 198918 198918 198918 633528 633528 633528
633528 45642 45642 45642 45642
[ 16] 45861 45861 45861 45861 36645 36645 36645 36645 23207 0 0 0 0 0 0
[ 32] 67557 67557 67557 67557 67557 67557 67557 67557 67557 35039 35039 35039 35039 35039
35039 35039 35039
[ 48] 26978 23033 23013 23013 19088 19088 19088 19088 67557 67557 67557 67557 67557
67557 67557 67557
[ 64] 23191 0 0 0 0 0 0 0
```

- Was all that struct re-organisation really worthwhile ?
- Why is there a huge un-used hole in my structure ?
- Discovered some amazing berkelydb allocations at startup eg.
 - Lots of poorly packed structures / classes too ...
 - Good number of false positives: 'opendir' eg.

Component based everything ... (UNO)

- ▼ UNO: an excellent / the best component model
- ▼ Unfortunately: over-reaches in scope ...
 - ▼ Automatic **bindings** for **scripting languages**
 - ▼ Interfaces for **core application** development
 - ▼ Hyper-granular (per method) **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
 - ▼ Poor technical decisions

What we're working on first:

- Enabling parallel contribution
 - Layout for dialogs
 - Help → Wiki
 - Easier build
- Making the code suck less
 - Unit test / better QA test-tool
 - Less redundant cruft
 - Improved performance, less memory & fewer leaks
- Smaller download footprint; not 160Mb x 50 for mirrors
- And lots of individual fixes

Conclusions ...

Corporate Ownership Aggregation ...

- ▶ Prepare for the marketing onslaught
 - ▶ well meaning people do dumb things every day
 - ▶ tell your pointy haired boss: worth a try
- ▶ Before contributing, check more than the license
- ▶ Resist the sirens of corporate collusion
 - ▶ ingratitude towards excessive corporate dominance
- ▶ Positively contribute to more open alternatives
 - ▶ MariaDB – LGPL [client library project](#)
 - ▶ systemd, LibreOffice, IcedTea, gtk+, vote with your feet.
- ▶ Read up on the [legal issues](#)
- ▶ Check out the [GNOME guidelines](#)

LibreOffice conclusions ...

- ▶ Aspire to a big, diverse, vigorous, vendor neutral community
 - ▶ cf. the Linux Kernel's healthy mob
- ▶ Strike a **dis-proportionate** blow for freedom
 - ▶ LibreOffice will have a **hundred million users**
 - ▶ We need to make sure they get excited about, instead of disappointed with Free Software
 - ▶ Try an '**Easy Hack**' - get stuck to the tar-baby ...
- ▶ Thanks for all the plumbers' help, support & advice
 - ▶ freedesktop.org / Xorg guys / Kernel mafia

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