The infrastructure of the LibreOffice project

- Florian Effenberger
  Alexander Werner
- LibreOffice Conference
  Milan, September 26, 2013

Why do we need infrastructure?

We are a "virtual" project

- the community ("the project")
  - **millions of users** worldwide
  - **hundreds of contributors** around the globe
  - active **24** hours a day, **7** days a week
  - across all **languages** and **time zones**
    → "the project **never sleeps**"
- the foundation
  - the board is located **worldwide**
  - we have **no physical office**

Infrastructure…

- …is best when you **don't recognize** it
- …is required **24x7**
- …has to be dynamic scaleable and **stable**
- …**enables** the community to work
- …is the **framework** for community work

Infrastructure provides the playground where all the magic happens

Where we came from

- **before** LibreOffice: proprietary framework
  - mainly tied to one IT service provider
  - many integrated services plus some external
  - **infrastructure defined** community structure
- with LibreOffice: **open and free** system
  - **community defines** infrastructure
  - we had the "luxury" of **starting from zero**

Our credo

Our infrastructure development reflects the **freedom, openness and transparency** that TDF stands for

What we are cooking with

Three-tier approach

- **dedicated servers** rented by TDF
  - mailing lists, websites, wiki, online help, MirrorBrain, rsync, CMS, WebDAV, OTRS, blog planet, Torrent tracker, AskBot, forum, Gerrit, extensions & templates – **most of things seen by the public**
  - **6 physical servers** and up to **29 virtual machines**
  - Intel Xeon **Quad-Core**, **32 GB RAM**, **SATA** disks, **Gbit**
- dedicated servers hosted by **third parties**
  - buildbots and tinderboxes for **compiling**, but also e.g. **Pootle**
- **external services** managed by others
  - BugZilla, developer mailing lists, blog, conference calls
  - **libreoffice.org accounts** for mail, Jabber/XMPP and SIP

Some numbers

- team of **11** administrators from **4** countries
- **6** TDF servers, up to **29** VMs, plus external stuff
- monthly **~8 TB** traffic
  w/o download & development
- **112** mirrors worldwide for downloads
- **127** mailing lists in **35** languages
- **18,000+** subscribers on non-developer lists
- in total **~170.000** e-mails in non-developer lists
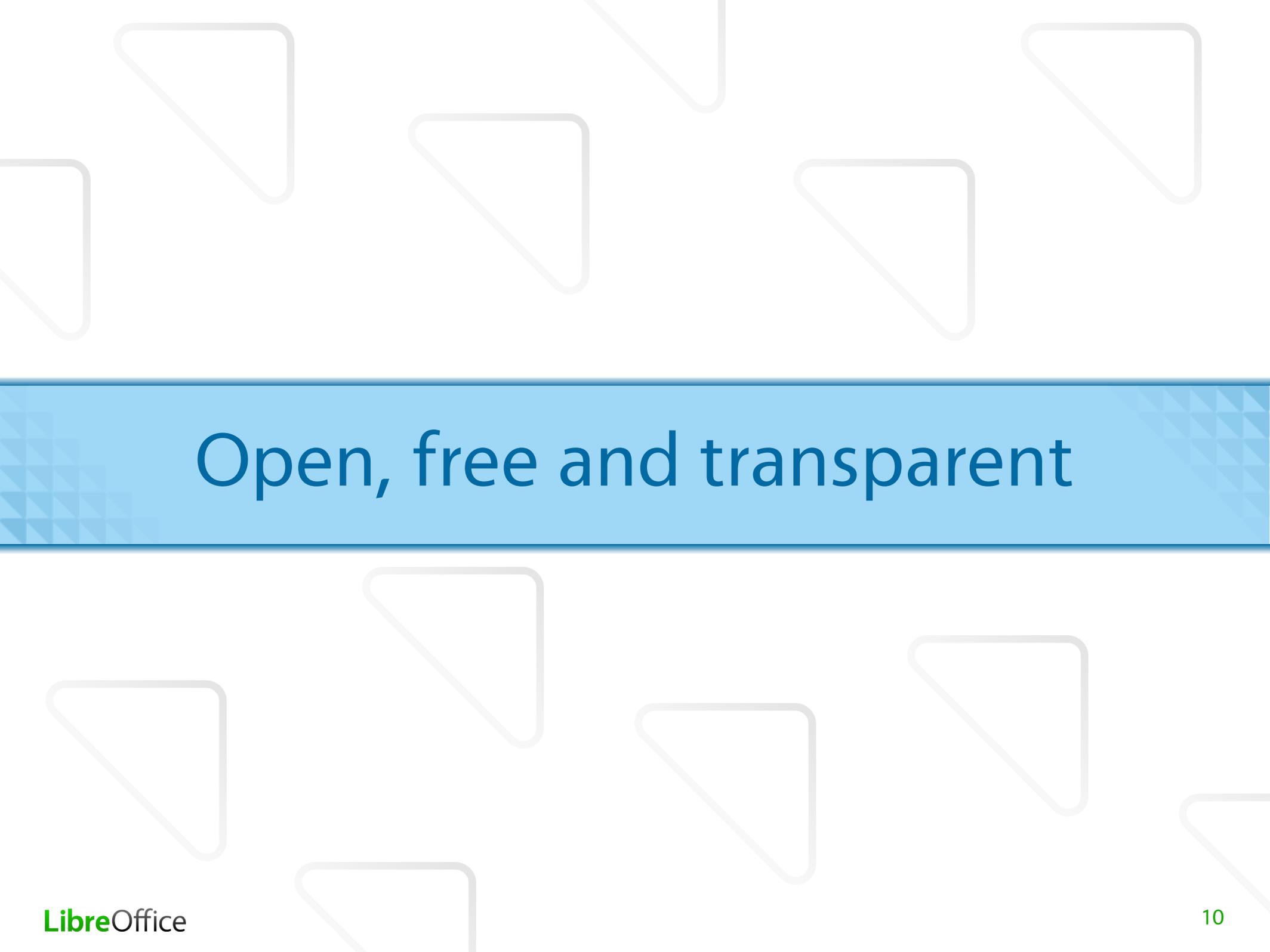
Open, free and transparent

LibreOffice

Including the community

- all **knowledge** is within the community
  - regular admin **meetings** in real life and via phone
  - professional **documentation** using Sphinx
- everything is based on **free software**
- we **contribute back** to other projects
  - **PyMIME** mailing list filter (stripping HTML)
  - translations and bug reports for **mlmmj**
  - **SaltStack** modules
  - **MirrorBrain** IPv6 repository also used by others

Building on reliable free software

- **Ubuntu** 12.04 LTS as base system, w/ LVM+RAID
- **Postfix**, **Dovecot**, amavisd, SpamAssassin, ClamAV
- **Mailman** and **mlmmj** for mailing lists
- **MySQL** and **PostgreSQL** as databases
- **Apache/Nginx** with **PHP**, plus **Silverstripe**
- **MirrorBrain** and **Mirmon** for mirror management
- **MediaWiki, Etherpad**, Planet, Plone
- internally: Nagios, OTRS, BackupPC, ownCloud
- all services are **IPv6**-enabled

MirrorBrain - the mirror mastermind

- **C++** Apache module, **Perl** and **Python** scripts
- **database of mirrors** w/ respective (**priority**)
- selection of best download server based on **GeoIP**
- files are automatically **scanned, indexed and hashed**
- **GPG signatures** are verified automatically
- log files for **download statistics** are generated
- **three-tier** mirroring system
  - 12 fastest hosts are **push mirrors** with "bit flip"
  - 47 **staging mirrors** with "bit flip"
  - 53 regular mirrors

Glimpse into the future

LibreOffice

Migration plans

- evaluating **Debian** 7.0 as second base OS
- migrating MySQL to **PostgreSQL**
  for **replication** and **performance** reasons
- converting from LXC to **KVM**
  for better **guest isolation**
  and virtualization of Windows hosts
- migrating from Apache to **Nginx** for performance
- evaluating Amazon EC2 **cloud** for certain hosts
- → plan to migrate to **fully virtualized services** later
  - HA with **failover** and loadbalancing
  - DRBD, Gluster

Central deployment with SaltStack

- infra needs to keep pace with community
  → organic **growth** to provide services in time
- right now, **handcrafted scripts** for initial setup
- in the future: **SaltStack** as deployment solution
  - one "Salt **master**" server with configuration repository
  - servers and virtual machines are "Salt **minions**"
  - deployment and regular **configuration** update
  - centrally managed **user database** (without LDAP)
  - base system already done, setup in **~5 minutes**
    → optionally including automatic creation of **cloud** instances

Join us, it's fun!

Communication channels

- **#tdf-infra** at irc.freenode.net
- **website@global.libreoffice.org** list
  subscribe with website+subscribe@global.libreoffice.org
- own wiki category at
  https://wiki.documentfoundation.org/Infra
- mail the admin team at
  hostmaster@documentfoundation.org

Thank you for your attention!

Any questions?
- E-Mail: floeff@documentfoundation.org
            alex@documentfoundation.org
- Twitter: @floeff / @alxwerner