Sympa usage survey: results analysis

1 Survey conditions and returns

The survey was available from September 17 to November 5, 2015. It consisted of 4 sections:

- knowledge of the respondents' structure (area of activity, geographical area, etc.)
- technical environment (essentially software bricks)
- volumes and consumption of resources
- exploitation the features of Sympa
- comments and opinions on other dissemination tools

157 unique responses were provided to the questionnaire, of which 122 were completed (i.e. all questions were answered).

The surveyed population was the set of subscribers to the lists of the community of Sympa administrators (sympa-fr and sympa-users), i.e. 994 people. The return rate was excellent (over 15%). This is a convoluted way of saying that the community is great and always ready to support development work. Thank you!

With such a large data population, the law of large numbers is respected. We can therefore extrapolate using descriptive statistics tools.

2 User population

2.1 Estimate of the total population

The survey collected users' preferred methods of installation. We now know that the servers are installed as follows:

- *61% from sources,* downloadable from the Sympa website
- *at 39% using packages from* Linux distributions

We know the number of websites downloading Sympa from Sympa in one year - ignoring multiple downloads from one website: 2150. These 2150 websites correspond to a minimum of 2150 servers installed from sources that make up 61% of the population.

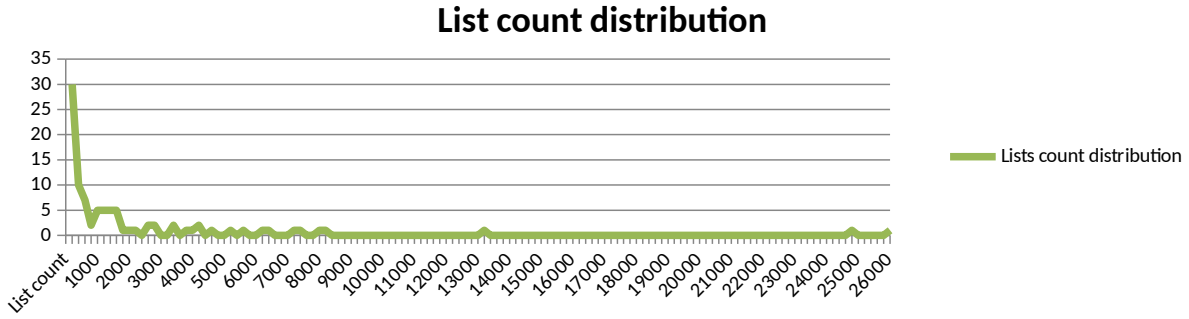
The simple rule of three enables us to estimate a minimum of **3,500 servers worldwide**.

2.2 Definition of the representative data of the population

Furthermore, we collected information on the numbers of users and subscriptions, as well as messages sent and received.

A general trend for all digital data is *long-tail distribution*: There are a large number of responses for small values and a decreasing number of responses as the values increase.

Below is an example of the number of lists per server:



The median of the distribution is 425 lists. Half of the servers therefore have less than 425 lists. Yet some servers have up to several tens of thousands of lists. Although not representative of the majority, they are certainly one facet of Sympa's use and cannot be left out.

With these type of non-normal distributions, it is not possible to use the arithmetic mean to determine a central trend, because it is too highly influenced by the extreme values. By way of illustration, the arithmetic mean of the number of lists per server in the population of respondents is more than 1500. One quick look at the graph above will immediately cast doubt on the relevance of this value as being representative of the majority. The standard deviation is 3,572. In other words, the arithmetic mean is not meaningful, except to reveal the extreme diversity of installed Sympa servers.

The **geometric mean** is far more suitable to this type of distribution. In fact, one usually observes that the geometric mean is fairly close to the median value.

It is therefore this indicator that we have chosen to characterize the average values of the population.

A table showing, for each measured value, the cumulative total for the population, the arithmetic mean, the median and the geometric mean, is attached. We also specified the standard deviation so that no one can complain that we haven't used the arithmetic mean.

2.3 The average Sympa

Once this choice has been made we can provide a snapshot of the average Sympa server (rounded values for easier reading, the exact values are attached) (we're not making them up):

Hardware resources:

- 2 CPU cores
- 4 GB of RAM

Use:

- only 1 list domain
- 370 lists

- 10,000 users
- 35,000 subscriptions
- 300 incoming messages each day
- 5000 outgoing messages each day

2.4 Extrapolated figures

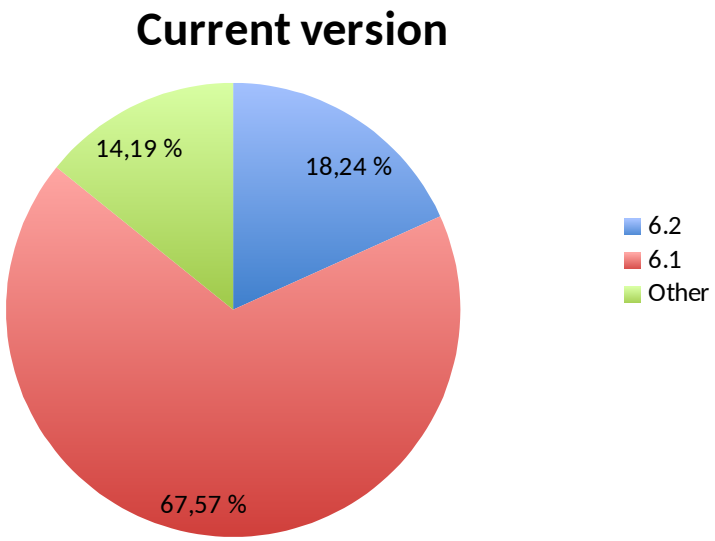
As we have reached the law of large numbers, we can legitimately use the values obtained and the estimate of the number of servers to extrapolate values for the global population. We can therefore estimate that there are, **globally:**

- **3,500 list** servers
- **1.3 million mailing lists**
- **35 million unique** end-users
- 1**30 million subscriptions**

We were really surprised at this.35 million is ridiculous compared to Facebook, but it's still 0.5% of the world’s population.Okay.End of self-satisfaction.

Moreover, we can also say that Sympa servers **receive 1.2 million emails and send 17 million each day.**

2.5 Sympa version

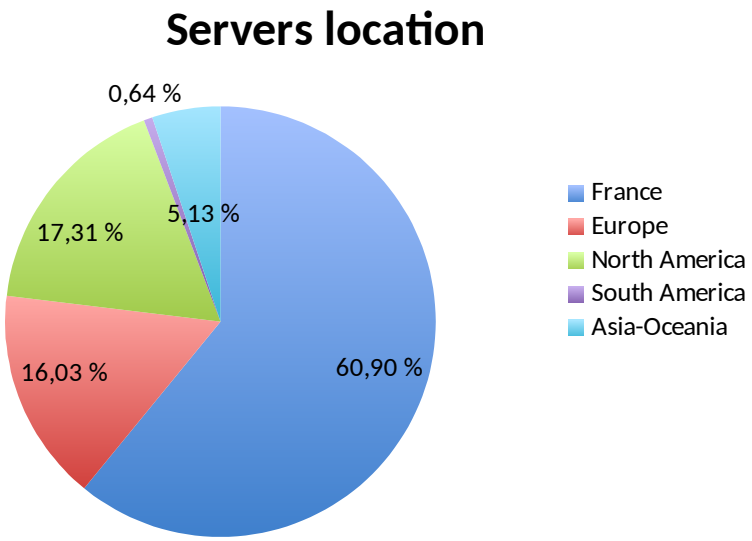


Sympa is a piece of software that lives on servers without disturbing anyone. It is therefore common for the version to expire without the administrator noticing. The results of this survey show, however, that administrators are becoming more and more scrupulous about updates.

While 68% of the servers were still on version 6.1 in November 2015 (6 months after the release of version 6.2), only 14% of the servers had a lower version, and 18% were already on 6.2.

Of those not running version 6.2, 85% said they wanted to install version 6.2 in 2016.

2.6 Geographic location



Most of the servers are located in France (61% of the population). This may be due to two factors:

- On the one hand, although the developer community is largely international, *the software is supported by RENATER, a French organization*; moral support from RENATER obviously encouraged wide deployment of the product in the hexagon.

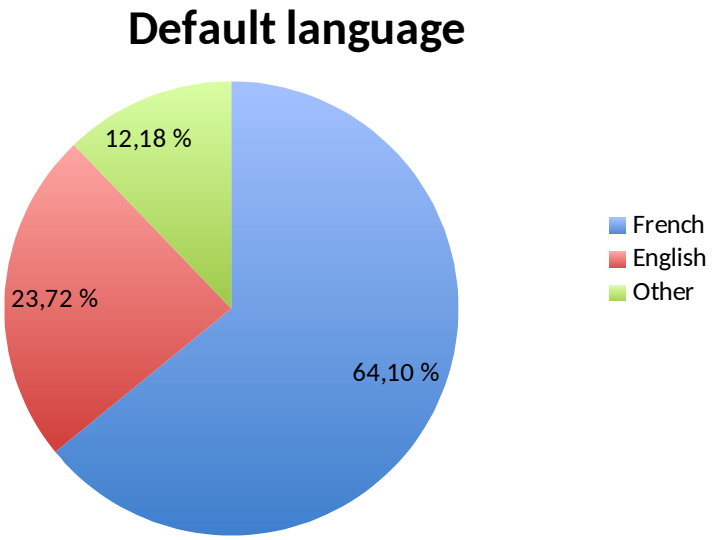
- there may be a survey population effect: the French user community has been there for a long time and is large; *it represents half of the respondents, which may have led to an exaggeration of the results.*

- Lastly, the only means of advertising the survey was sending emails to the lists of sympa-users and sympa-fr, if other communities exist outside these lists, they probably have not seen the survey.

However, it remains quite obvious that, if we focus on results outside of France, North America and Europe are the major centres where Sympa is deployed, with 17 and 16% of servers respectively. Asia and Oceania (5%) and South America (1%) are almost non-existent. No installations have been identified in Africa.

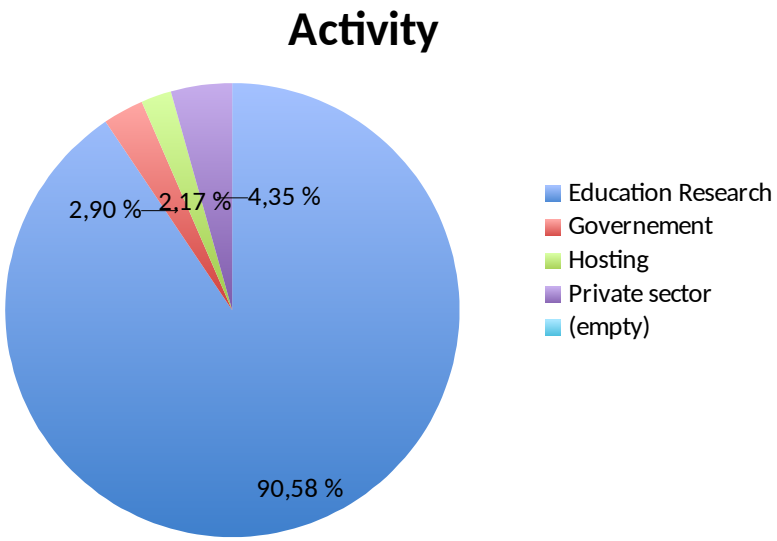
2.7 Language

Unsurprisingly, given the geographic distribution of the servers, 64% use French as their main



language; 24% use English. **Although Sympa has been correctly translated into 12 other languages, these represent only 12% of the servers**. It can be concluded that the English version of the software is used in many countries.

2.8 Field of activity



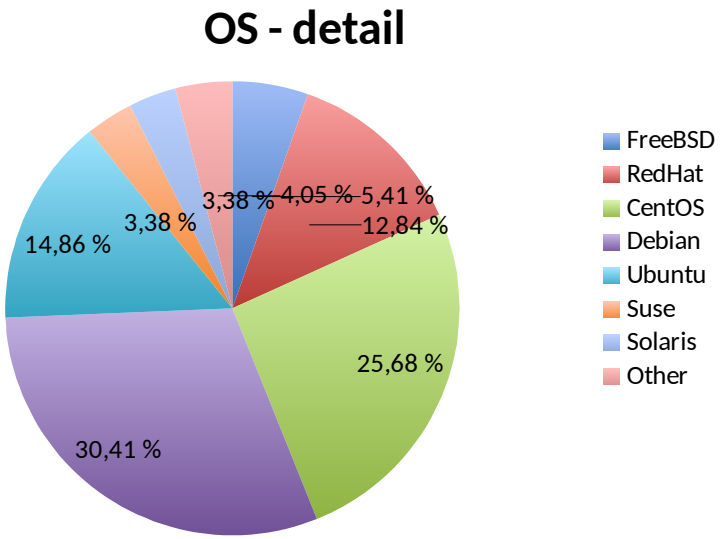
Sympa is mainly used in higher education and research institutions (83% of servers). This is not surprising as the software originated in this community and is particularly suitable for their needs.

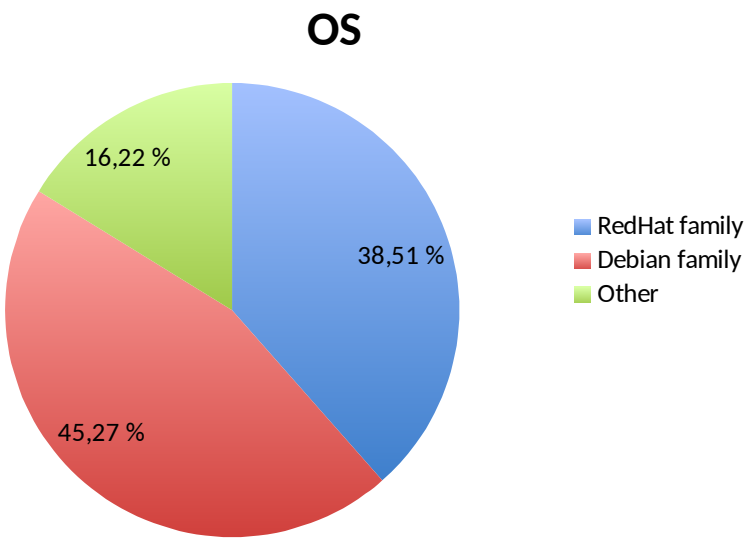
The second largest users are, surprisingly enough, associations (8%). While it is normal that

associations, which are usually underfunded, use free software, it is curious that other areas have not started using Sympa, which has many advantages in terms of integration into information systems (provisioning from data sources, in-depth management of access rights, multiple authentication methods, etc.).

3 Technical environment

3.1 Operating System

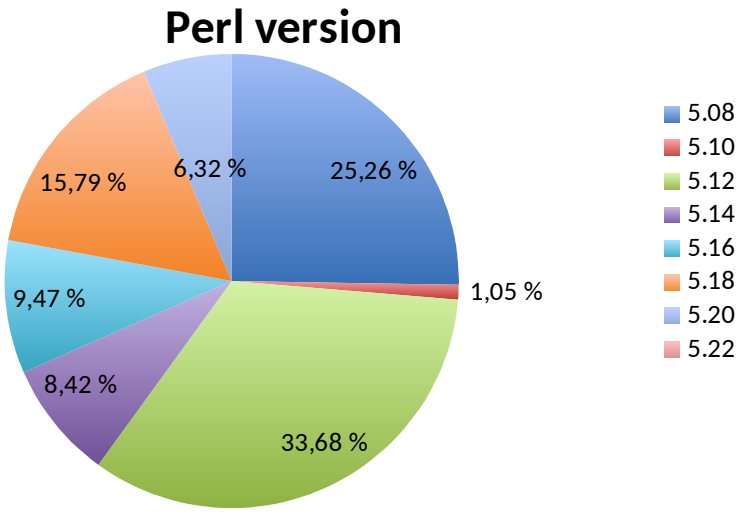




The big players are the two main families of Linux: Debian and RedHat, including all distros, with 45 and 39% of servers respectively. This leaves a total of 16% for other systems: FreeBSD, Solaris, Gentoo and Suse mainly.

It should be noted that packaging is not a determining factor. Debian, Gentoo and FreeBSD are the only distros to integrate Sympa into their official repositories. It is therefore not Sympa that determines the operating system used.

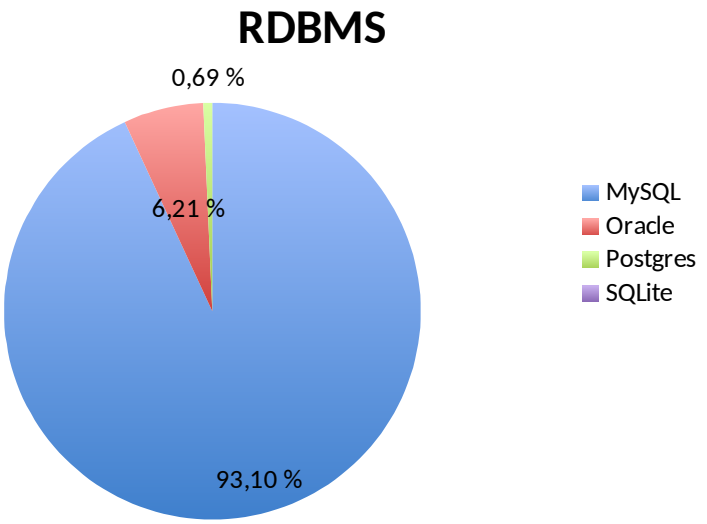
3.2 Version of Perl



Sympa being 95% Perl code, it is important to know which version is installed and the features available.

A fair amount of uniformity is observed, all versions from 5.08 to 5.22 being represented. In particular, **16% of servers are using 5.08, which means that we still cannot use 5.10 typed code**. Sympa's prerequisites in terms of the Perl version will not change for a year or two.

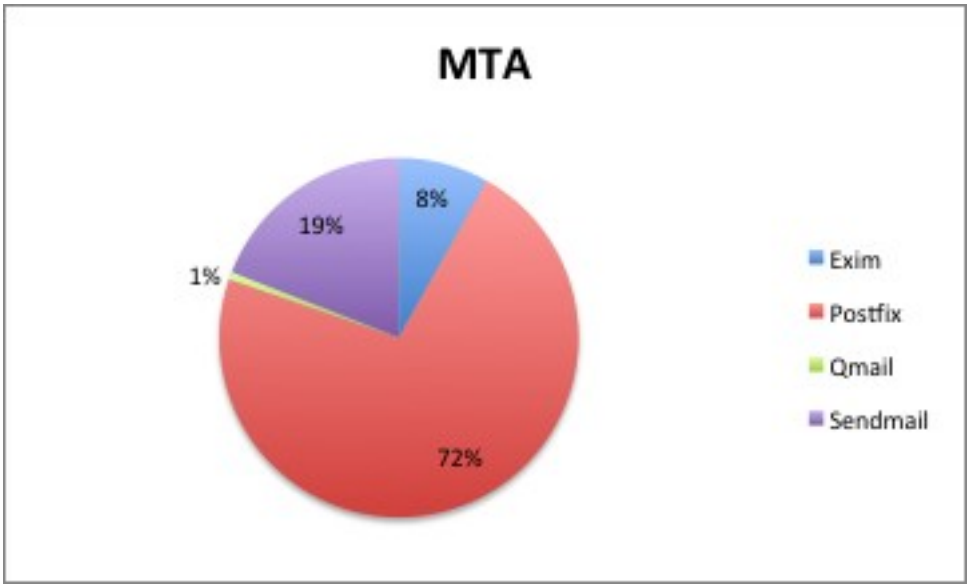
3.3 DBMS



With a banana republic score (92% of servers), **MySQL is the big winner.** This is normal; **it has long been the only DBMS for which database updating is automatic**. Three other DBMS (SQLite, PostgreSQL and Oracle) now also have automatic updating. The big loser is Sybase:

nobody uses it. We will **definitely abandon support for Sybase in the next version of Sympa.**

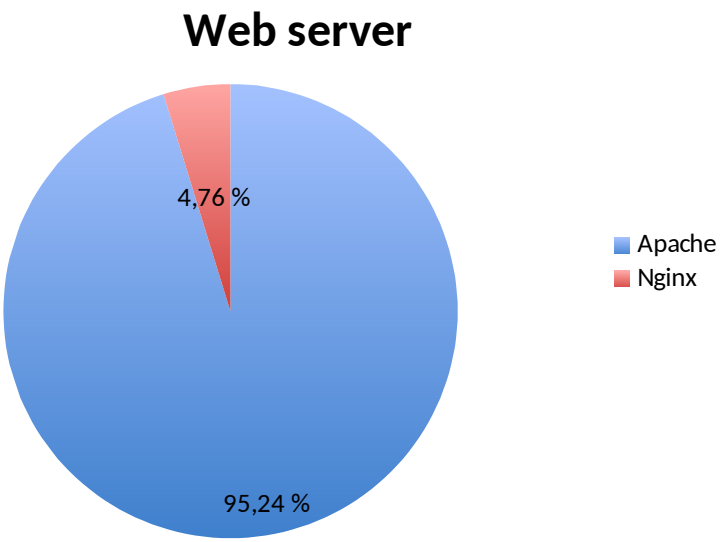
3.4 MTA



Postfix is the leader by far, with 72% of Sympa's facilities running it. Sendmail is not yet dead, however, with 19% of servers. Exim is already well developed at 8%, Qmail remains anecdotal at 1%.

The mail environment therefore remains very heterogeneous. There is therefore no question about Sympa using the specialized functions of one MTA rather than another.

3.5 Web server



Apache.

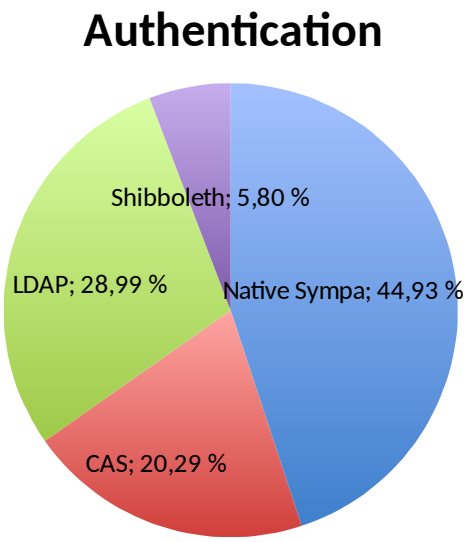
95% of the servers are running Apache. Which is an expected result as Sympa uses Fast CGI mode,

which is extremely easy to install with Apache, but far more complex with its only competitor, Nginx and its 5% of brave users.

This broad dominance may encourage us to seek, as has been suggested, to more closely integrate Sympa into Apache. This would simplify the installation process.

It is possible that the use of Perl web frameworks such as Dancer will change this trend, as they enable the engine (FastCGI, modperl, etc.) responsible for the HTTP response to be abstracted. This is an ongoing idea among the authors of the software, to be confirmed / refuted in the future.

3.6 Authentication



Of the 4 possible authentication systems (CAS, LDAP, federated and native Sympa), only federated authentication remains marginal, with 8% of respondents using it.

In 45% of cases, it is simply native Sympa authentication that is used. It is likely that, in these cases, Sympa is used autonomously and independently of other tools.

The massive use of CAS and LDAP (almost 50% of servers) reflects Sympa's place as a business tool, since third-party authentication requires up-to-date repositories which are only found in companies .

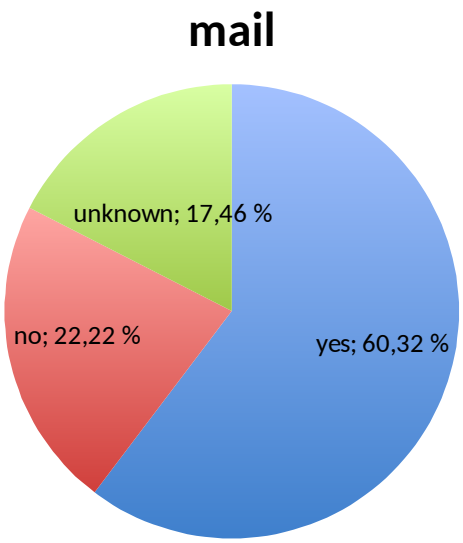
4 Use the Sympa’s features

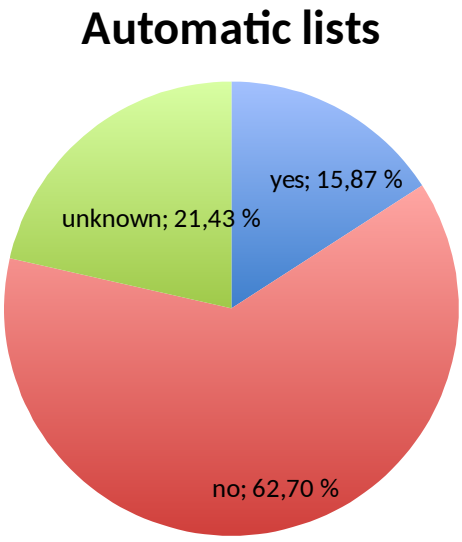
In the survey we tried to find out which functions of Sympa are used and which are not. Respondents were asked if they used the following functions:

- data sources for feeding lists
- the SOAP interface
- the web interface

- the email interface
- the command line
- list families
- authorization scenarios
- web templates
- email templates
- shared documents

4.1 Email interface and automatic lists: the big unknowns

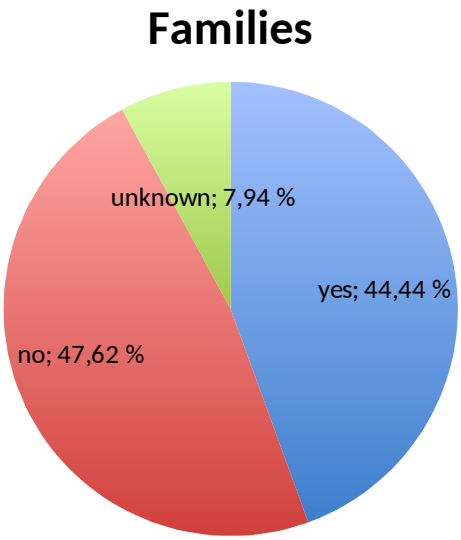


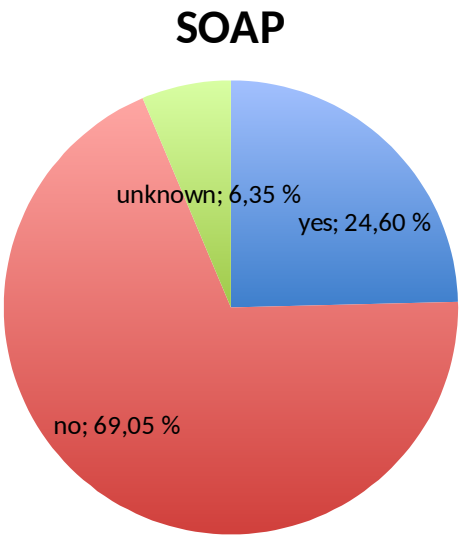


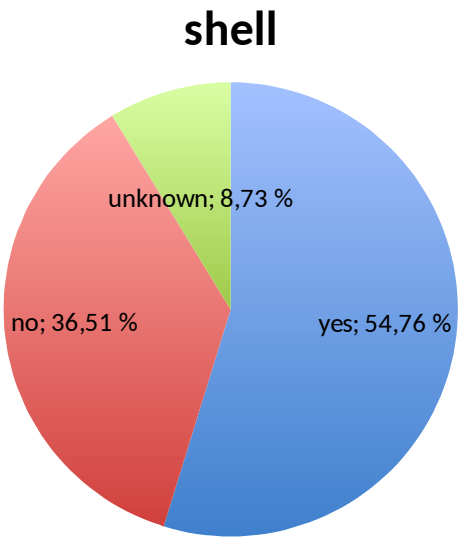
While most of the features of Sympa are well known to users, people are still fairly unfamiliar with the email interface (unknown to 18% of respondents) and especially automatic lists, which 22% of respondents admit they only discovered at the time of the survey.

While families are well represented (44% of the instances use them), automatic lists remain the poor relative of Sympa, with only 16% of servers implementing them. This is hardly surprising, considering the complexity for end-users with slightly dated versions of Sympa. It should be noted that, from version 6.2, automatic lists can be handled using clicks in Sympa's web interface, making things far easier.

4.2 The unloved ones: list families, SOAP and command line





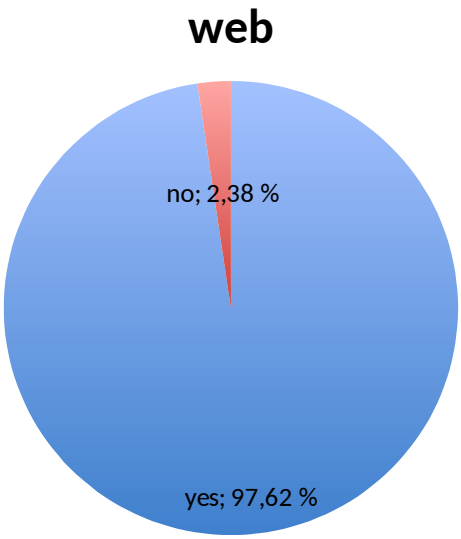


Three other functions are less well known: list families (unknown by 8% of respondents), the SOAP interface (6% of respondents) and the command line (9% of respondents).

Despite the relative lack of awareness, the command line interface is fairly widely used (55% of respondents). This interface should therefore be expanded and strengthened.

It is striking to note that, while most respondents were aware of the SOAP interface, it is still not widely used; this especially as an expansion of this interface was one of the requirements expressed by users (see the last part of this document). It is likely that the REST interface, once in place, will find an audience more easily.

4.3 Sympa's interfaces



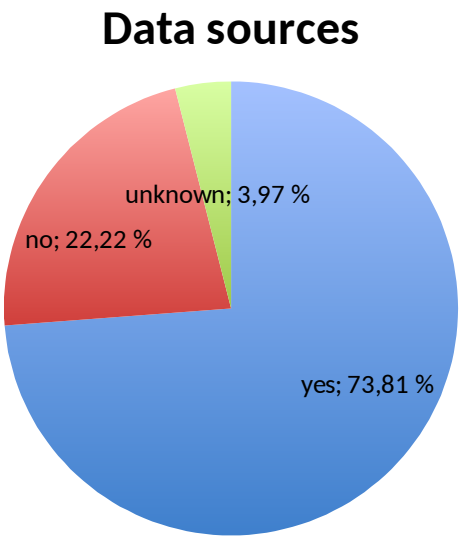
It is very useful and deserves to be better known.

While 98% of respondents use the web interface, only 60% use the email interface, 55% the command line and 25% the SOAP interface.

We should perhaps say: 25% are using the SOAP interface "already"! This means that Sympa integrates well with information system that are remotely controlled or which remotely control other applications.

Shared documents are still used by 45% of instances, despite the many existing document sharing solutions available. This function is therefore still useful as a file-sharing tool, probably due to the complementarity of its communication (sending messages) and document sharing (shared documents) functions within a centrally managed group (the list) and managed by end users.

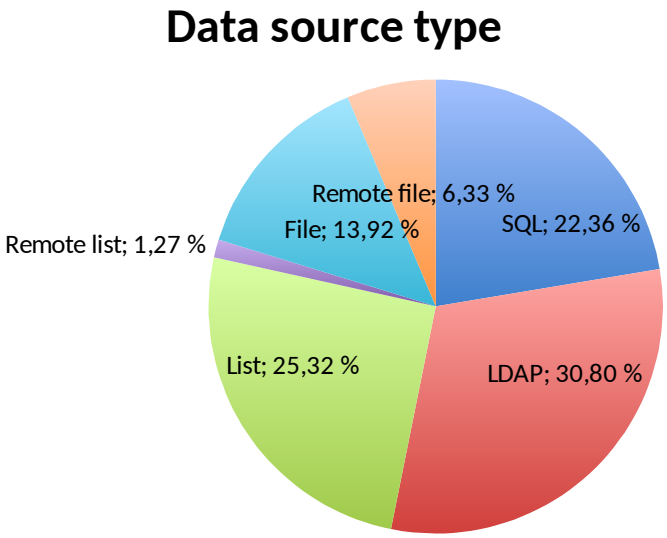
4.4 The star: data sources



The principle feature is still data sources for feeding lists. 74% of instances use it, delegating the management of lists to the information system in this way.

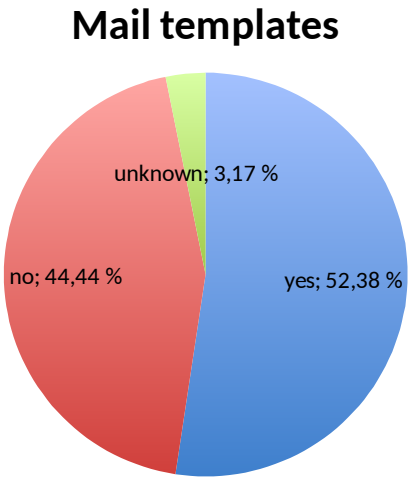
It is interesting to note in this regard that most types of data sources are exploited, indicating the heterogeneity of the instances and their environments.

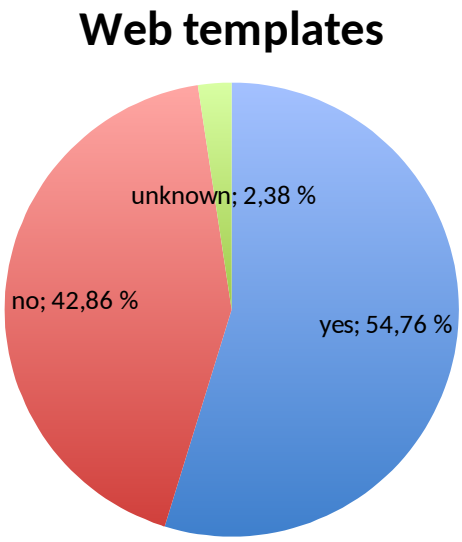
As with centralized authentication, the use of data sources denotes integration into an information system. The good representation of the inclusion of lists, a data source very easily usable by non-technical users, indicates that the delegation of responsibilities in Sympa is well used.

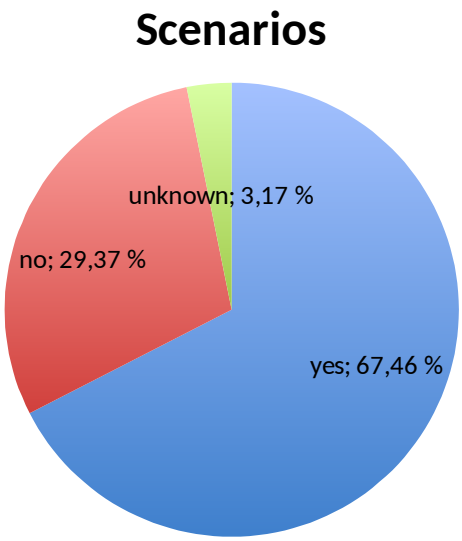


4.5 Customization

As some of the many customizable elements of Sympa, we asked users if they had modified web or email templates, or scenarios.







Contrary to our expectations, we discovered that they are often customized as more than the half of the respondents had already adapted them to their needs. This figure even rose to 68% for scenarios, which we consider to be one of the key features of Sympa.

The flexibility of use and suitability of Sympa for various needs is therefore one of the main advantages of the software, which is also confirmed by the comments provided at the end of the survey.

4.6 Past and present

For 36% of respondents, Sympa has replaced another tool, mostly Mailman or Majordomo, mostly because Majordomo was at the end of its life and Mailman had not (yet) developed information system integration capabilities.

5 Comments

We now come to the analysis of the comments. There, we leave behind statistical objectivity for a more global analysis. I apologise in advance for any approximations or errors of judgement that may appear in the following text. Despite this possible flaw, it seems relevant to point out that we included a comments section in the survey and will therefore have to answer the questions posed.

5.1 What do you like about Sympa?

- the data sources
- automation (families and automatic lists) that negate the use of scripts.
- Scaling up
- The flexibility of the software, its many customizable features
- the delegation of tasks (listmaster, owners, moderators and subscribers), the autonomy it provides and therefore (even if it was not written for this) the load-lightening that this represents for administrators.
- Authorization scenarios and the granularity provided to control processes.
- The fact that Sympa is open source
- The ability to use virtual hosts
- Archives
- The fact that Sympa is written in Perl :) a language which administrators often know.
- the web interface
- The dynamism of the project (support and development) and the ability to challenge the authors (who always welcome comments as they are great people)
- Shibboleth authentication

People like Sympa for what they can do with it, since all the qualities noted are broadly used according to the survey.

5.2 How do we make a better Sympa?

- Curiously, the most requested feature is **automated reply** for lists during periods of absence (requested 4 times!),
- and a more ergonomic and richer email sending feature. It is also asked that message sending be done in the context of **modernized archives**. Which could be summarised as being like the Mailman hyperkitty project, a web forum, with social network-type features, which sends a message to the list when one writes to it.

- In the same spirit, we are being asked to improve the ergonomics of shared documents.

- To facilitate the integration of Sympa with other services, especially messaging tools, by extending its **web services.** We are being asked to extend the SOAP service to all Sympa functions and to create a REST service.It was even requested that Sympa's web interface only be a client of a REST service.

- Simpler management of **families** and their expansion to all lists, which would enable better control over the created lists and mass updating of the configurations of all lists.

- We were also asked to look at the **life cycles of** the lists; one recurring request refers to the ability to automatically close lists based on a pre-defined date or period of inactivity.

- In terms **of software administration**, it has been suggested to accelerate the transfer of the source code to the packages. This is in line with other requests which are more about simplifying the installation and update process.In the same vein, it was strongly suggested that we simplify the application logs and offer better supervision capabilities.It was also suggested that we remove our home configuration framework and replace it with a standardized framework such as XML or YAML.We are being asked to provide a modular configuration including the ability to enable / disable features in Sympa to make it easier to secure and save resources.Also relating to security, we have been asked for a global blacklist.Users would also like to be able to manage lists across the entire server ignoring the robot.This would be part of a global dashboard, also requested, which would oversee all error lists and statistics at the server level.Lastly, the listmasters could be provisioned from a data source.

- **Scaling** of Sympa, if recognized as an asset, could be further improved, including the generalization of load balancing capabilities and improving responsiveness when the number of lists becomes very large.

- Several requests regarding **messaging**: aliases for list names, automatic compression of image attachments. It has also been suggested that we introduce permitted message sizes depending on the sender. Similarly, it was proposed that Sympa could automatically insert an email address in CC for the messages in a list. Lastly, someone mentioned the integration of features such as mailchimp - including the insertion of cookies into emails. Forget this last point. We will not be doing it.

- Few people mentioned improvements to the Sympa documentation, which was one of our fears. But we remain committed to working on this aspect.

- **User** management was also mentioned. Of course we are talking about a long-running story: the management of multiple email addresses for a single user. We also want to be able to search and delete all of a user's subscriptions.

- Last but not least: we were asked for a coffee machine. That's good: we needed one too!

5.3 How are we going to handle your requests?

In short, our software development plans roughly cover what you are asking for.

We have factored the code, the objective being to reach a sufficient level of modularity such that all functionalities of Sympa are accessible on all of its interfaces: web, email, command line, SOAP

and REST (yes, we have already started on REST). This goes hand in hand with other developments, such as homogenization of configurations, which will certainly be based on YAML. Lastly, we want to move towards an SaaS configuration, which would enable people to set all configuration elements online. This implies simplifying the management of families, with good reason, and providing global control over the lists in order to carry out actions which, in shell, require sed executions.Specifically, this will involve generalizing the concept of the family to all lists.

Improvement of the web interface will continue with:

- A dashboard including statistics on the server. We already save them anyway, they just have to be displayed.
- Forum-type organization for archives. This would provide the ability to use Sympa using the archives only, if desired. We already store all the information you can see in hyperkitty, we simply have to exploit it with the required Foundation libraries.

We also want to enable Sympa to be multi-channel. Understood: we want Sympa to disseminate information using several methods of communication: email, but also web services, SMS, Atom, etc. Not by implementing them itself, but by enabling it to connect to services doing it for it.

That’s for the future.

In the immediate future, it should be noted that you also asked us for things that the latest version of Sympa can already do:

- Use .incl files for subscribers, not just owners and moderators (https://www.sympa.org/manual/parameters-data-sources#member_include)
- Include white lists; possible thanks to the plugins (https://www.sympa.org/manual/templates_plugins) from Steve Shipway (http://www.steveshipway.org/software/f_sympa.html)
- A space for sharing documents: that's been around for a long time, but we're going to improve it and, yes, one day you'll have drag and drop as well.
- A collaborative space (text editing, document sharing): Sympa can’t do it, but we have set up coupling with Dokuwiki precisely for this reason (https://www.sympa.org/templates_plugins/dokuwiki_plugin)
- Tracking receipt of messages (https://www.sympa.org/manual/bounces#message_tracking)
- Warnings when moderated messages are approved / rejected
- Subscriber searches by name or address: or subscriber searching in the web interface
- Statistics: yes we save them all in the database, and we already present a some of them in the web interface of each list.
- A responsive web interface: this was the major update in 6.2.Thanks to the University of Auckland for providing the first version!
- recovery of attributes other than the email by LDAP: now possible in LDAP and SQL (https://www.sympa.org/manual/parameters-others#include_sql_ca)
- unsubscribe with one click from an email: this is possible with the unsubscribe URL (https://

www.sympa.org/manual/message-handling#unsubscription_url).Note that this is not just one click: the user must confirm it by clicking on a link received by email.

- TLS support: Ayet since version 6.2!

- remove the address from the exclusion list without resubmitting it: yes, we got rid of this chore.In a list using data sources, you now have an "Exclusion" section that lets you remove a subscriber from the exclusion list.

Conclusions: update to 6.2 as you had planned.:-)

5.4 Potential replacements

We also asked you what you would replace Sympa with. The only software mentioned were:

- Mailman for a lot of good reasons, the most often mentioned being archiving, more modern and functional than in Sympa. It was also reported that configuring Sympa lists is discouraging for end users (note that opn can simplify this with edit_list.conf). Lastly, a mailman argument is the whitelist function (now possible in Sympa thanks to Steve's plugin).

- Groupserver: provides cross-administration functions, i.e. simultaneous modification of several lists, as well as a REST interface. We have also been criticized for poor spam management, but this specific criticism was based on lack of knowledge of the authorization scenarios.

- Listserv, mentioned once "because it’s the tool that the management had in its structure before arriving here".

5.5 Q&A

Some comments were questions rather than suggestions. Here are the answers:

- why is there a new version every 2 weeks?

    - *We estimate Sympa's installed base to be 3500 servers. This implies very heterogeneous environments. So we always get bug reports after releases of major versions, which we try to resolve quickly. We sometimes add lightweight features as well, and Soji always has ideas to improve the code. Hence the regular releases of Sympa.*

- When will you move to github, saperlipopette?

    - *Actually, there are already Sympa repositories on Github, but none are maintained by us.The code base of Sympa 7.0 (the one currently being worked on) is in a git repository in Sourcesup.We did not put it on github because it will likely be changed a lot in the future and therefore does not constitute a reliable code base.Once refactoring is satisfactory, we will give you the green light to work on github.The reference code will stay on Sourcesup because we have a set of continuous integration tools on it.But we will accept pull requests from Github.*

- Could not you provide a Debian repository to avoid the delays inherent with the Debian freeze?

    - *Two things: contributors to the Debian package (Emmanuel Bouthenot in particular) do a great job and we thank them. If they deem it useful for such a repository to exist in the sympa.org domain, we will gladly make it available to them. But otherwise, no.*

- Why did you remove the check boxes for accepting or rejecting moderated messages en mass?

    - *It seemed like a good idea at the time. Well, obviously not. We will give them back.*

- How do you do X?

    - *X is in the doc! Oh... Wait... Okay, we'll re-work the doc.*

6 Lastly

This is neither a suggestion for an improvement nor a question, but we found this remark in the free comments. We will reprint it verbatim:

“The Sympa community and developers are very constructive to work with (which makes using it and working on it a pleasure). "

We agree. That's why we like developing this software.

That's all, thank you again for taking the time to answer this survey and letting us know what the community thinks, you are the only ones worth following.

And now to work.

7 Attachments

Cumulative results for the numerical values of servers, as well as the various indicators calculated using the data collected.

	Total	Average	Standard deviation	Median	Geometric mean
Num. of cores	432	3	4	2	2
RAM	5,089	39	359	4	4
Number of virtual hosts	669	5	18	1	2
Number of lists	212,553	1,586	3,572	400	344
Number of users	6,687,755	51,444	103,721	8,368	7,559
Number of subscriptions	29,990,909	232,488	472,274	42,677	32,281
Incoming emails / day	263,656	3,296	9,860	350	300
Outgoing emails / day	2,502,095	32,922	51,874	10,000	4,646