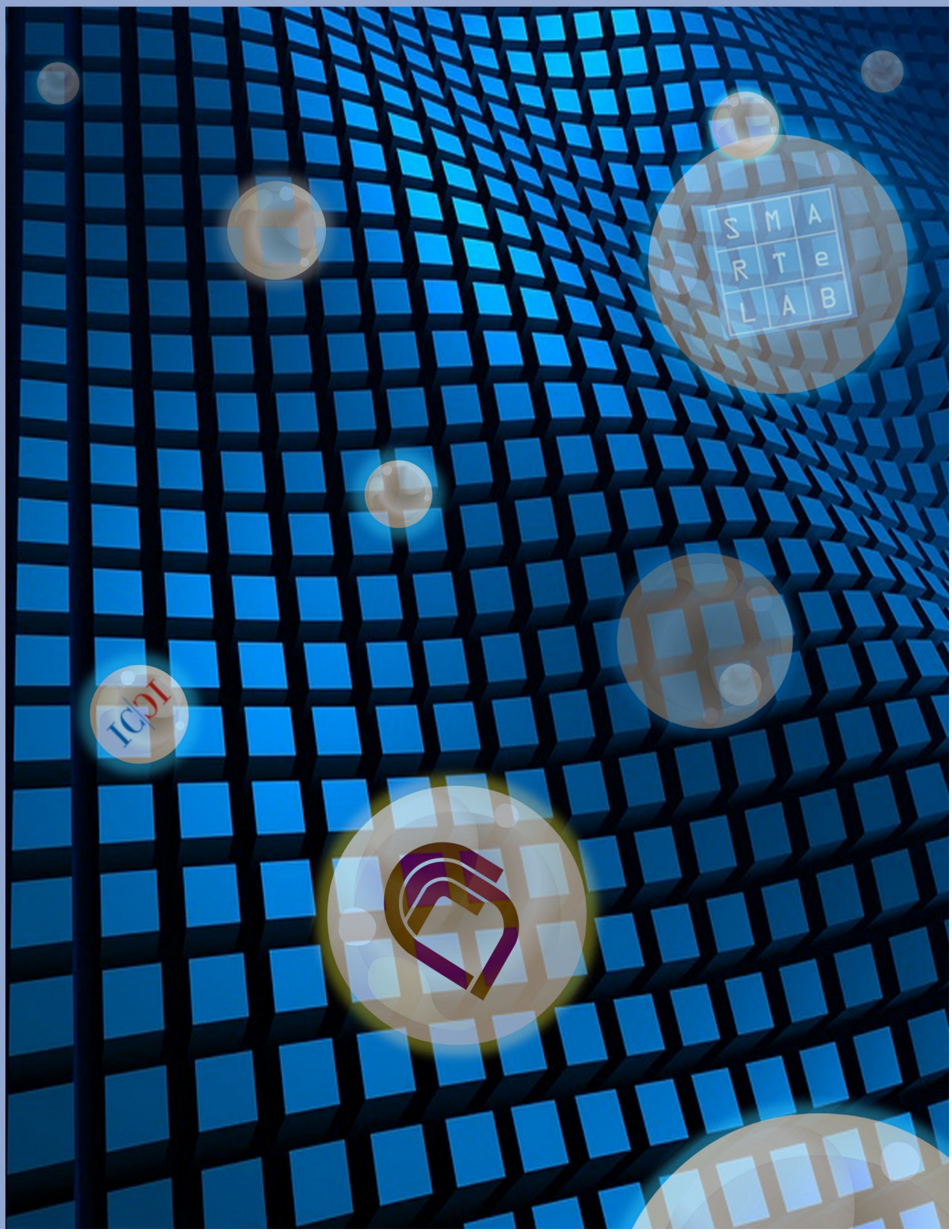
SMART eLAB

VOLUME 8- ANNO 2016



ISSN 2282 - 2259

SMART eLAB

SOMMARIO
Vol. 8, 2016

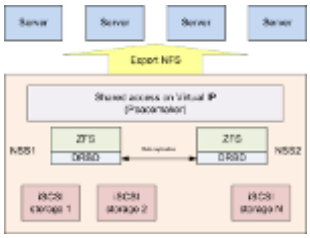
Articoli

Smart e-Lab: http://smart-elab.mlib.ic.cnr.it
A peer-reviewed online resource, published by the Istituto di Cristallografia (CNR-IC)
EDITORS-IN-CHIEF : Michele Saviano, Augusto Pifferi - ASSOCIATED EDITOR : Guido Righini
GRAPHIC DESIGN : Claudio Ricci - EDITORIAL ASSISTANT : Caterina Chiarella
CNR - Istituto di Cristallografia, Strada Provinciale 35/d, I-00015 Monterotondo, Italy

Networking

Setup of a redundant network storage system. A legacy approach.†

Andrea Lora,[a] Giuseppe Nantista,[a] and Augusto Pifferi.[a]



In order to provide an affordable network storage for general purpose servers, we setup a system that is able to share NFS resources avoiding single point of failure. This architecture is now used as a storage for our email server. The goal is achieved using opensource software and different vendor storages.

Keywords: HA storage, DRBD, NFS, ZFS, iSCSI.

1 INTRODUCTION

When providing a general purpose server the first goal to achieve is to avoid that a single fault on a component of the system can cause loss of data or, at least, long time of unavailability of the system. Server that offer those services must be designed to work on more physical machines, located in different IDCs, running different instances of the service, but offering same data to clients. This can be obtained by replicating every single server and assuring that in every moment data contained in the first server are the same of those contained in other.

However this approach can limit the overall scalability of the system, so the only way to get both replicated service and mirrored storage is to separate the application and the storage layer. The main advantages of this structure are two: high scalability, in fact it is possible to add other storage in order to create a clustered pool of resources; single maintenance is needed, no matter how many services use the infrastructure; servers are easier to manage, because data are kept outside.

1.1 Overview

This paper will propose a possible implementation of this scenario, using Linux as the operating system for the network storage servers, iSCSI[1] as the network storage protocol used to access different SANs from storage servers, DRBD as the replication engine, Corosync and Pacemaker as the clustering and resource management daemons used by network storage system servers, ZFS as the local file system and NFS[2–7] as the shared access protocol. This kind of cluster is often referred as *Active Passive - Shared Nothing* architecture.

The described infrastructure is shown in Fig. 1

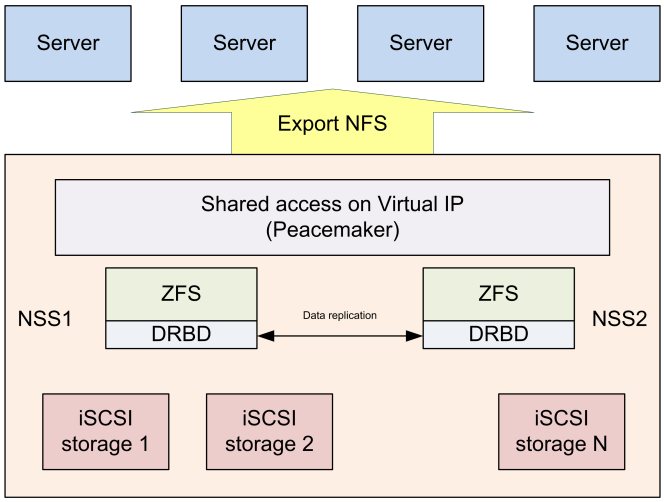


Fig. 1 Overview of the architecture

1.2 Hardware

Implementing an active/passive cluster means we have two (or more) machines configured to do the same job. Only one of the machines is actually serving the resources, while the others are put in stand-by, ready to take over the service in case the primary get some kind of failures. We'll cover details of our configuration in section 2.

1.3 Network Layer

Nodes need to talk each other to keep them synched and serving data to clients. We also need a floating IP which will be assigned to the active server. The configuration of the virtual IP is explained in the section 7.

1.4 Block storage layer

There are three main storage networking standard for linking data storage facilities, those are iSCSI, Fibre Channel and Infiniband. Both Fibre Channel and Infiniband require dedicated

[a] *Istituto di Cristallografia, C.N.R., via Salaria km 29,300 - Monterotondo, Italia*

† Rapporto tecnico IC-RM 2016/02 protocollato in data 14/04/2016 n. 658

network hardware. The only protocol that allows the use of generic hardware such as network switches and cards is iSCSI. It also permits opensource implementation of the protocol stack.

We used three storage systems, a Syneto Unified Storage, a self assembled storage based on Illumos operating system, running Nexentastor and a HP Lefthand P4000. Then we configured pairs on LUN of the same dimension on two storages, and accessed those LUN from the network storage servers.

1.5 Replication system layer

To avoid SPOFs we need to be sure that all the data is stored in (at least) two location. This will avoid data loss in case of failure of the block storage. We chose DRBD as replication engine. We'll cover details about it in section 4.

1.6 File system layer

Since DRBD is seen by the OS as a block device we needed to choose a file system suitable for hosting data. ZFS was our choice, due to its excellent records in data storage and powerful capabilities. Implementation of ZFS on Linux requires some efforts in the configuration due the fact it's not tightly integrated in the OS. We'll cover ZFS details in section 5.

1.7 Shared access layer

After creating the file system we need some means to access it from remote machines. Industry standard to allow access of remote file system is NFS. There was no reason to go against it, in order to follow a legacy approach. We'll cover NFS details in section 5.

1.8 Orchestration

Due to the clustered nature of this stack cluster/orchestration tools were needed. Pacemaker and Corosync were used as cluster suite. Although it's possible to command corosync/-pacemaker through the *crm* command line utility, we used the Linux Cluster Management Console (LCMC) to do most of the work. The presence of a GUI allow to get at a glance overview of cluster status. Additional customization was made to activate STONITH mechanisms. We'll cover cluster details in section 7.

1.9 Monitoring

The fault tolerance capabilities of this kind of stack don't mean that we simply forget it because it works. We implemented some custom script to take analytics of the performances of cluster nodes and setup some alerts accordingly. We'll cover the details is section 8.

1.10 Backup

Cluster nodes are always seen a single entity. The fact that data is redundant doesn't mean that it's guaranteed to be safe. A simple *rm –rf* can (and will) destroy data on the active and passive node at the same time. The passive is not the backup of the active, therefore backup strategies need to be implemented to avoid data loss. We'll cover details about backup in section 8.

2 Hardware

The cluster consist in 3 Virtual Machines. Two are the active/passive nodes of the NFS server, one is a very light instance acting as a quorum node. The fat nodes are equipped with 4 virtual CPU, 8GB of RAM, a bunch of disk space (16GB). The thin node is just 1 CPU, 512GB of RAM and 4GB of disk space.

The virtual machines are spread between two ESX systems. One is a fully fledged VMware Cluster hosted on an HP Blade System, the other is an HP G5 Server. The HP Blade System has attached a HP Lefthand P4000 as storage system, while the G5 has a Syneto Unified Storage. Both the storages offer a 1TB raw LUN accessible via ISCSI from the OS of the virtual machines, and are also in charge of hosting the data of the disks of the Virtual Machines we used.

Network side we configured 3 different networks and VLANs accordingly, one for NFS share, one for Replication/Heartbeat, one for ISCSI. At present the VMware Cluster hosts one fat node and the thin (quorum) node, while the ESX standalone G5 hosts the second fat note.

Please note that this configuration isn't truly SPOF free, because in case of complete shutdown of the VMware cluster, the last node will not be able to take over the job due to the lost quorum. Simply move the quorum machine elsewhere and you'll reach full SPOF free configuration.

After the configuration of the virtual machines we installed Ubuntu 12.04 Server on each server. The Ubuntu choice was due the fact that PPA repositories were available for the zfsonlinux project, an essential part for the system, and the LTS support from Canonical.

3 Block Storage Layer

After installing the OS on the fat nodes we enabled access to the ISCSI resources installing the iscsi-initiator-utils via apt-get, then we proceed with the discover of target:

```
# iscsiadm -m discovery -t st -p ip_of_portal
```

After that we should log in to the portal

```
# iscsiadm -m node - target_name -p ip_of_portal -l
```

And configure automatic login at boot

```
# iscsiadm -m node -T target_name -p ip_of_portal --op update -n node.startup -v automatic
```

Through the use of dmesg or with fdisk –l we should now see the presence of a new block device into the OS. In our case our LUNs were 1TB large. It's important that the LUNs size is the same for both the fat nodes. Failure to do so may cause problems with the replication.

4 Replication System Layer

Due to our design requirements we needed that the LUNs were perfectly synchronized, every time, any time. The software that allow to reach this goal was the Distributed Replicated Block Device DRBD. We can understand it as network based RAID-1 (mirror). DRBD offers asynchronous (A mode), semi-synchronous (B mode) and synchronous (C mode) replication. A DRBD device is seen from the OS as a standard block device, which can be manipulated like any other: from the OS point of view a DRBD device is just another hard disk. DRBD consists in two components: one runs as a daemon, it uses a couple of tcp port to ensure bidirectional communication, the other is a kernel loadable module. DRBD can be used without orchestration tools, but that won't be useful to achieve HA because you need to manually control primary/secondary promotion.

In order to install DRBD we first install the package trough apt-get

```
# apt-get install drbd8-utils
```

After the installation of DRBD we need to create a resource for it. A resource is the block device DRBD will offer to the OS. In the resource definition there will be present the physical block device where DRBD will write. Take into account that DNS names and IP must be configured accordingly with your network topology. We used ip in the configuration, but we also edited our /etc/hosts to provide coherent information. Rely on DNS to resolve the hostname is not advisable due the possible failure of the service.

```
resource r0 {
      protocol        B;

      startup {
              degr-wfc-timeout        0;
      }

      net {
              max-epoch-size  8000;
              max-buffers     8000;
              unplug-watermark        8000;
              cram-hmac-alg   sha1;
              shared-secret   my_shared_secret;
      }

      disk {
              on-io-error     detach;
              no-disk-barrier;
              no-disk-flushes;
      }

      syncer {
              rate    25M;
               al-extents      3389;
               csums-alg       md5;
               verify-alg      md5;
               c-max-rate      100M;
```

```
        c-min-rate      10M;
    }

    on nfs-1 {
        device          /dev/drbd0;
        disk            /dev/sdb;
        flexible-meta-disk      internal;
        address         10.10.73.26:7788;
    }
    on nfs-2 {
        device          /dev/drbd0;
        disk            /dev/sdb;
        flexible-meta-disk      internal;
        address         10.10.73.27:7788;
    }
```

In this example we define a resource called r0 synched semi-synchronously (B type). It consist in two hard disk being synched on the machines (nfs-1 and nfs-2). On both nodes the disk used is /dev/sdb and the block device exposed by DRBD is /dev/drbd0. We define an internal flexible meta-disk, storing metadata of DRBD inside the replication block device. We also tuned some parameters about the syncer, in particular we set the maximum transfer rate in 25 MB/s, and we choose md5 as the checksum and verify algorithm. Others parameters are in the net section and those needs to be tuned to your particular workload.

We then initialize the meta-disk area. We need to do this only on one node.

```
[root@nfs-01 etc]# drbdadm create-md repdata
About to create a new drbd meta data block on /dev/sdb.
 . ==> This might destroy existing data! <==
Do you want to proceed? [need to type 'yes' to confirm] yes
Creating meta data... initialising activity log NOT initialized bitmap (256 KB) New drbd meta
data block sucessfully created.
```

After that we can start DRBD on both nodes. Just type

```
service drbd start
```

If we check DRBD status (*drbd-overview* command is a beautiful wrapper that show us info) we'll see that both nodes are secondary, and not synched.

```
[root@nfs-01 /root]# drbd-overview
0:r0  Connected Secondary/Secondary Inconsistent/Inconsistent A r-----
```

We shall promote one of the nodes as primary (we will use nfs-01)

```
[root@nfs-01 /root]# drbdadm -- --overwrite-data-of-peer primary r0
0:r0  Synctarget Primary/Secondary Inconsistent/Inconsistent A r-----
```

After a while a drbd-overview will show

```
[root@nfs-01 /root]# drbd-overview
0:r0  Connected Primary/Secondary UpToDate/UpToDate A r-----
```

This will informs us that the LUNs are synched. Take into account that we didn't specify a runlevel for autostart of DRBD, and that's normal. In fact the DRBD resource will be managed by the cluster suite. More informations on clustering tools is available on section VII.

5 File System Layer

We now reached our first goal: we have two different LUNs synched. Every write operation we do on nfs-01 will be replicated to the nfs-02 node. But that's just a block device, we need a file system over that to be actually useful. Our file system of choice was ZFS. We wrote a small white paper about why ZFS is the best filesystem to be used for data storage at this time, so we won't speak about its features, but we'll focus about how to integrate it in our system. Although it is possible to use ZFS in user space trough FUSE there are serious drawbacks in performances so the suggested operating method utilizes kernel loadable modules. Installation is simplified due the presence of ppa repository for Ubuntu. Installation steps are as follow:

```
[root@nfs-01 /root]# apt-get install python-software-properties
[root@nfs-01 /root]# add-apt-repository ppa:zfs-native/stable
[root@nfs-01 /root]# apt-get update
[root@nfs-01 /root]# apt-get install ubuntu-zfs
```

After executing this commands we are ready to use ZFS on Linux. We can load the module through modprobe, but if the OS detects a ZFS volume on the disks will try to autoload the module automatically. There's a small caveat on this: since DRBD will replicate the volume in Active/Passive mode the secondary machine will actually see a ZFS disk pool and will try to load the module. This will fail due the read only DRBD property on the secondary and will cause some problems in the stack. We must exclude ZFS module from the autoload list. The scripts that manage the cluster resources will be on charge of loading/unloading it. To avoid the auto load of the module we simply add the following lines on the /etc/modprobe.d/blacklist.conf

```
blacklist zfs
install zfs /bin/false
```

The first inhibits the autoload of the module, the second specify that ZFS will need to be loaded only through an *insmod* command and not through *modprobe*.

Usual rules for ZFS administration apply. Since the block device underlying ZFS is replicated, a ZFS command will propagate to the second node, so this actions must only be done on the active ZFS node.

We start creating a ZFS pool and referencing it with the correct block device.

```
zpool create tank /dev/drbd0
```

Since /dev/drbd0 is our replicated block device it is the correct one to be passed to zpool as an argument. *Tank* is a common name for a zpool, but it's your choice. Creating a pool automatically creates a ZFS filesystem with the same name, but in order to ease administration it's good practice to create nested ZFS filesystem inside the pool.

```
zfs create tank/share1
```

Here we command ZFS to create a nested filesystem inside the tank pool, and we call it share1. That's particularly useful because you can change some ZFS attributes (compression for example) on the single filesystem, rather than the pool. It also helps you managing different snapshots policy pool based.

One of the most important tune to make to ZFS is to disable the atime option of the filesystem. With atime enabled every time you access a file the filesystem updated the metadata regarding its last access timestamp. This lead to a write operation for every file read, quickly degrading the performances of the system. You can completely disable the atime property with zfs set atime=off or use a lighter timestamp called relatime. With *zfs set atime=relatime* you are telling the filesystem that you want an update of the last access timestamp, but just one every 24 hours.

Please note that we avoid to use ZFS capabilities of auto-exporting filesystems via NFS. We decided to manually export them via the usual linux tools. That's because we need more fine grained control on the events: the clustering tools are in charge of orchestrating everything. Since we blacklist the ZFS module in the modprobe the clustering tools must me take charge of loading it, we'll cover this kind of details in section 7.4.

Take into account that ZFS is not a clustered filesystem: this mean you can't have ZFS mounted on both the server at the same time serving the same data.

6 Share Access Layer

In order to access the replicated resource we just created from remote machines we choose to use the nfs server. NFS was one of the most used protocols to share files between machines in unix environment: it requires few dependencies and the protocol is old but reliable. The NFS server can be installed through apt-get install nfs-kernel-server. If you are using NFSv3 configuration is limited to the /etc/exports file, where you can add the directory you want to export and the ACL and options for them. NFSv4 need an additional component, named *idmap*, which is a daemon that map remote uid to local ones. We don't need to configure the /etc/exports at this time nor we need to enable nfs server at startup. The orchestration tools are in charge to dynamically load the kernel modules and update the exports list.

The system is designed to offer high availability and transparent migration from one node to the other. NFS clients whose operations were on the wire will have trouble reconnecting timely. We need to adjust two parameters to lower the grace time of NFS and smooth the failover. There parameters are nfsv4leasetime and nfsv4gracetime, both tunable in /proc/fs/nfsd

7 Orchestration

All the orchestration is done by corosync/pacemaker. Corosync in the software that manages cluster membership of the nodes, determining quorum and promoting/demoting nodes. Pacemaker is the resource manager that starts and stops the various services and provide fencing agents for unresponsive hosts. Corosync and pacemaker are often called together "cluster tools". Configuring them was usually done through config files, but a brilliant project called "Linux cluster manager console" allows to use a GUI to install and configure the cluster tools. You'll find the relevant configuration files in Appendix. The use of LCMC is pretty straightforward. Pacemaker actually performs the operations through agents. These agents are often bash scripts with certain exit codes. Two main kind of agents exists: LSB and OCF. LSB are considered legacy modules provided to maintain compatibility, OCF is the new format and is becoming pretty popular and more manageable.

7.1 Corosync / Pacemaker

Corosync configuration in its simplest form is limited to declaration of IP address, network card to be used to cluster communications and multicast address.

Pacemaker allow us to configure certain kind of relationship between resources such as “start A before B”. This is a core feature to load the various components: we need a bottom up approach when starting resources, and vice-versa when stopping them. LCMC shows a nice view with the relationship of the resources, and allows to configure the resource agents we need for getting the job done.

Caveat: a cluster should be configured to start resources only when the appropriate fencing agents are started to avoid the risk of split brain situations which can be very difficult to recover without data loss.

7.2 Fencing agent

Fencing agents are in charge of killing a node in case it doesn't answer to keepalives: usually they use STONITH approach via IPMI or remote power switches. Since we are in a virtual environment STONITH is done via VMware api calls. In case of VMware VCenter there is a well know agent: *stonith_external/vcenter*. If you are using an unlicensed version of ESXi you can rely on ssh to poweroff the guests. You can find an example of ESXi fencing agent in the Appendix. If fencing agents fail to start for any reason, your cluster won't be able to start other resources. For debugging or devel purpose you can override through the parameter *stonith_enabled*. Please note that operating a cluster without stonith agents can lead to split brain situations and consequential data loss.

7.3 Block device agent

The block device agent in this configuration is *ocf_drbd*. It's a master/slave agent that ensures no two nodes are running the same drbd resource as primary. The only parameter this agent takes is the name of drbd resource. In our case it is r0. All the following agents run on primary node.

7.4 Filesystem agent

After the block device appears in the operating systems we need to mount the filesystem that it contains: in our case ZFS. We haven't found an ocf zfs agent, we ended writing our own. We need to implement the action start / stop / check. For start action what we need to do is manually load the ZFS module that we blacklisted before and running a zpool import. For stop action we need to zpool export the pool and remove the module from ram. For check action we can rely on zpool list. At this stage the agent takes two parameters: device to be mounted and mount point. The zpool name is hardcoded as tank.

7.5 NFS Server Agent

Nothing much to say here. A predefined ocf agent named *nfsserver* manage the task of running and stopping the NFS kernel server. As parameter it takes the script to start the nfs daemon.

7.6 Exports agent

After the NFS server is started we can describe various exports directory. An OCF agent named *exportfs* is available for such task. It takes as parameters the client ACLs, local share name, a unique *fsid* within the cluster for that share and the export options, if any.

7.7 Virtual IP agent

Clients don't connect to the real IP of the servers, because they have no clue which NFS server is primary. We define a last resource, a Virtual IP assigned to the primary server. The OCF agent is called *IPAddr2*, and the only mandatory parameter is the IP address you want to use. You can also specify a netmask if needed. This IP is what the client will use to access the nfs share.

7.8 First Start

At first start corosync form the cluster, assign one of the nodes as DC (Designated Controller) and check for pacemaker configuration. Then fencing agents are started and checked for status. If they are ok the next agent kicks in, starting drbd on both nodes, promoting one to Primary and the other as secondary. Sync starts here if needed. Only on the primary node as soon block device is confirmed to be available, filesystem is mounted, nfs kernel started, directory exported and finally IP address assigned to the primary node. If any of the steps fails to complete pacemaker perform a rollback taking down the resources on the primary node until block device layer. Then it perform a role switch on the nodes: the primary is demoted to secondary and the secondary promoted to primary. After it tries to perform all start actions on the new primary.

7.9 Failover

If failover is requested pacemaker first tries to take down all the services on the primary node. If stopping the services is successful it performs role switch on the nodes and try to start all the resources to the other node, which is primary at this point. If during the stop of the services on the old primary a certain timeout is met, or if them fails with error the cluster tools proceed to fence the unresponsive server. Often this mean a complete poweroff/poweron cycle of the server.

8 Monitoring

Even if cluster can perform automatic failover is good practice to monitor the server to get instant insight of what is happening in the system. The items we want to monitor are the system resources of the servers, the status of pacemaker resources and some statistics about performance of nfs. We installed the zabbix agent throught apt-get to get the basic stats of the server and we also added several custom checks to enable the monitoring of the pacemaker resources. We wrote a simple bash script to check the various services and added the zabbix_agent.conf the relevant keys. You can find both the script and the zabbix keys in the Appendix.

9 Backup

While operating a cluster the administrator must take into account that the secondary server is the replacement, not the backup. Secondary server are there to achieve HA, but a number of things can go wrong. Due the semi-synchronous nature of this setup both block devices are actually mirroring each other, like a RAID-1. So we need other strategies to achieve solid backup.

9.1 Snapshots

When we faced the choice of the filesystem we agreed on one thing: filesystem snapshot is just too useful to not have it. And this influenced heavily our choice in favor of ZFS. Its copy-on-write mechanism allows to take snapshots in nearly no time. More important snapshots aren't really occupying space on disks: they are just thin copies. And they are easily accessible through the hidden .zfs directory where the admin can just inspect or copy the files. What we needed was a system that automates the snapshots of the filesystem and rotates the backup in a way that respect our policy. We found in zfs-auto-backup the solution of this problem. This handful scrIPt (even available through apt-get) allows to take snapshot while defining a retention policy. We keep 4 quarter snapshot, 24 hourly snapshot, 30 daily, 4 weekly, 12 monthly and 5 yearly. If the system becomes bloated by the yearly snapshots we can easily offload them as explained in the next section.

9.2 Filesystem send

The *zfs send* command allow us to stream the whole filesystem to stdout. The *zfs recv* command takes a filesystem stream from stdin and store it in the designed pool. The capability of sending a filesystem, or just the incremental snapshots of it, is another powerful feature of ZFS and it comes extremely handy for backup strategies. We can offload old snapshots just sending them over the network to another machine that is ZFS capable. Due the fact that zfs send/recv use standard descriptors we can just pipe through an ssh connection. An example of backup is

```
zfs send tank@snap1
       | ssh host2 zfs recv tank/backup
```

ZFS properties can be different on destination. For example we activate compression on the destination filesystem. Sending data out of the cluster solves the problem of backup. We use the local snapshot for common activities, such restoring deleted files or accessing old versions of them. If the cluster has problem, or if we need do some heavy I/O load like a search in the whole snapshots for a certain string we can offload the work on the backup server which normally doesn't require to be responsive. A forked version of zfs-auto-backup feature auto send of snapshot to a remote host.

10 Conclusion

The work described in this technical report was made in order to obtain a storage system for general purpose in high availability environment, assuming multiple and contemporary access to it. The result obtained required a big effort in deploying it, resulting in very high probability of human error when managing it, furthermore we accepted some trade-off that limit the overall system performance, resulting in slow file access. However, the conceptual effort and the experience have highlighted some of the strengths and weaknesses of this system. Scalability is delegated to the iSCSI layer, growing the LUN results in the capability to immediately grow the file system, and allows, thanks to the adoption of ZFS, automatic and cost-free snapshot mechanisms and consequently enables granular recovery of individual files starting from the selected snapshot. Future works aimed at exploiting the benefits of distributed and scalable storage systems, such as CEPH, continuing to enjoy benefits of a file system like ZFS in the backup system.

Glossary

ISCSI:	Internet Small Computer Systems Interface
DRBD:	Distributed Replicated Block Device
NFS:	Network File System
SPOF:	Single Point of Failure
ZFS:	Zettabyte File System
LUN:	Logical Unit Number (SCSI devices)
LCMC:	Linux Cluster Manager Console
STONITH:	Shoot the Other Node in the Head (computer clustering)
COW:	Copy-on-write
IPMI:	Intelligent Platform Management Interface

Appendix

All the configuration files needed in this project can be found at http://code.sra.mlib.cnr.it/andlor/nfs

References

1 J. Sartran, K. Meth, C. Sapuntzakis, C. M., Z. E., Internet Small Computer Systems Interface (iSCSI), Request for Comments: 3720.

2 S. M. Inc., Network. File System Protocol Specification (NFS), Request for Comments: 1094. March 1989.

3 B. Callaghan, B. Pawlowski, P. Staubach, NFS Version 3 Protocol Specification, Request for Comments: 1813. June 1995.

4 S. Shepler, B. Callaghan, D. Robinson, R. Thurlow, C. Beame, M. Eisler, D. Noveck, NFS version 4 Protocol, Request for Comments: 3530. April 2003.

5 A. Galloway, Things about zfs that nobody told you. jul 2011.

6 A. Traeger, E. Zadok, N. Joukov, C. P. Wright, A nine year study of file system and storage benchmarking, Trans. Storage 4 (2) (2008) 5:1–5:56. doi:10.1145/1367829.1367831.

7 Solaris Internals wiki, ZFS Best Practice Guide.

Networking

Accesso Wi-Fi con Autenticazione Federata.†

Luca Ianniello,*[a] Augusto Pifferi,[a]

In questo documento viene descritta l'attività di realizzazione di un sistema Captive Portal per l'accesso a internet con Autenticazione Federata. Il progetto ha per obiettivo quello di consentire agli ospiti del Campus dell'Area della Ricerca Roma1 (CNR), che già posseggono una Identità Digitale certificata, un accesso sicuro con le proprie credenziali senza l'obbligo di una nuova registrazione dei dati personali pur mantenendo comunque il pieno rispetto delle Acceptable Use Policy (AUP) imposte da GARR che richiedono esplicitamente la tracciabilità degli utenti. Sono indicate le specifiche di progetto, l'architettura, il software e i risultati ottenuti.
Keywords: Access Point, Hotspot, Captive Portal, Wireless, IDEM, Accesso Federato, SAML, simpleSamlPhp, Service Provider, Identity Provider.

1 Introduzione

Presso l'Area della Ricerca di Roma1 (CNR) è presente una infrastruttura Wireless a 2.4Ghz/5Ghz composta attualmente da 30 Access Points distribuiti presso la maggior parte degli edifici.

La configurazione dei singoli AP è gestita in maniera centralizzata da un unico controller, un Aruba network 3600 che comprende ed implementa la possibilità di utilizzare un Captive Portal per l'autenticazione degli utenti (fig. 1).

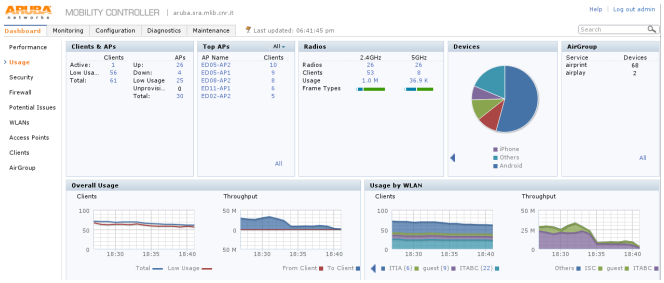


Fig. 1 Interfaccia di monitoring del controller Aruba3600

Ritenendo che il Captive Portal fornito dal controller fosse troppo restrittivo o non sufficientemente personalizzabile, oltre che oneroso in termini di licenze, si è optato per l'implementazione di software Open Source in grado di soddisfare le necessità dell'accesso a Internet via Wireless Autenticato e Autorizzato.

2 Il progetto

La tecnica "Captive Portal" consiste nel forzare un client http connesso ad una rete di telecomunicazioni ad accedere ad una speciale pagina web (di solito per l'autenticazione) prima di poter permettere la navigazione all'utente.[1]

In questo progetto, oltre alle oramai ben note dinamiche di accesso è stata prevista la possibilità di autenticare gli utenti attraverso credenziali gestite da altre Identità Federate, evitando completamente la prassi di registrazione dei dati personali. Per ottenere questo scopo occorre istituire un Service Provider e registrarlo presso la Federazione IDEM[2] del GARR.

3 Specifiche del progetto

Le specifiche del progetto sono:

- Uso di un software Open Source su piattaforma Linux;
- Implementazione del servizio sulla struttura hardware già esistente sfruttando l'Aruba3600;
- Implementazione di software per l'abilitazione alle funzioni di Service Provider SimpleSamlPHP;
- Registrazione del Service Provider presso la Federazione IDEM;
- Consentire la connessione esclusivamente verso l'Identity Provider appartenenti alla Federazione per l'ottenimento dell'Autenticazione;
- Impedire la navigazione ai client non autenticati;
- Consentire l'accesso ad account già connessi (multi-dispositivo cellulari/pc etc);
- Riepilogo dettagliato delle connessioni effettuate dall'account utente Sistemi di sicurezza per la salvaguardia del sistema operativo e dei dati;

[a] *Istituto di Cristallografia, C.N.R., via Salaria km 29.300, 00015 Monterotondo Italia.*

† Rapporto tecnico IC-RM 2016/05 protocollato in data 21/06/2016 n. 0001101

- Log degli accessi;
- Prassi di Logout fine sessione.

Il software Open Source “CoovaChilli”[3] per le sue funzionalità di Access Control Software è già stato utilizzato con successo in altri progetti dando prova di flessibilità e robustezza e pertanto è stato deciso di riutilizzarlo in questo. Grazie al supporto di SAML (Security Assertion Markup Language)[4] è stato possibile modificare le dinamiche di autenticazione per demandarle all’IdP di appartenenza dell’utente anziché al RADIUS previsto da CoovaChilli.

4 Realizzazione del CAPTIVE

Il Captive Portal con Service Provider è stato installato su una macchina virtuale del server HP con sistema operativo di base ESXi del Servizio Reti d’Area con le seguenti caratteristiche:

Modello	Virtual Machine ESXi-VM are guest
Processore	4vCPU
MemoriaRAM	Memory 3 GB
Dischi Fissi	Virtual Disk 16 GB
Scheda Video	VMware SVGA II Adapter
Scheda Rete	3 Virtual Network Adapter

Tabella 1 Risorse a disposizione nella Virtual Machine

Per poter usufruire di un AAI (Authentication and Authorization Infrastructure) che ha il compito di razionalizzare e semplificare i sistemi di autenticazione circa gli accessi ai servizi tra organizzazioni diverse, ci si avvale dell’utilizzo di SAML attraverso un apposito framework. Visto l’elevato livello di personalizzazione richiesto dal progetto si è preferito l’uso di SimpleSamlPHP sfruttando nel contempo la sua semplicità di implementazione.

Tramite SAML è possibile realizzare il Single Sign On istituzionale autenticando gli utenti localmente e autorizzandoli ad accedere alle risorse in base ad informazioni scambiate dalle organizzazioni in modo sicuro.

L’intera piattaforma è stata installata e configurata su Sistema Operativo GENTOO LINUX.[5]

L’implementazione del software RADIUS necessario in questa piattaforma è anch’esso di origine Open Source ed è FreeRadius.[6] Per la parte dei dati relativi alle credenziali d’accesso e gli attributi richiesti a supporto per il funzionamento del protocollo “AAA” utilizzeremo quelli forniti dagli appositi IdP tramite il SP creato con SimpleSamlPHP.

Una volta configurato SimpleSamlPHP per svolgere la funzione di Service Provider, come da manuale, sono stati generati i metadati necessari alla registrazione dello stesso presso IDEM.

La registrazione alla Federazione IDEM prevede due fasi: una prima fase di test durante la quale lo staff del GARR verifica e controlla la coerenza, la completezza e la corretta forma dei Metadati del SP o dell’IdP che si sta sottoscrivendo alla Federazione, per ottemperare a questa meticolosa fase il GARR mette a disposizione un portale dedicato al controllo dei Metadati prodotti facilitando e migliorando la stesura e la verifica degli stessi.[7] (Fig. 4)

Oltre ai Metadati necessari per l’accesso alla Federazione, tra i dati richiesti, occorre indicare all’interno della descrizione delle policy, gli attributi che saranno richiesti dal SP per l’accesso al servizio che si sta proponendo. (Fig. 5)

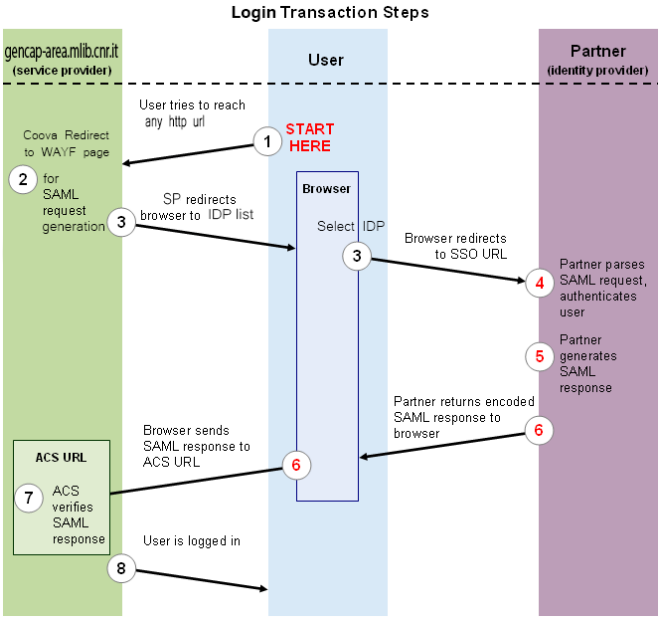


Fig. 2 Fasi di autenticazione e autorizzazione

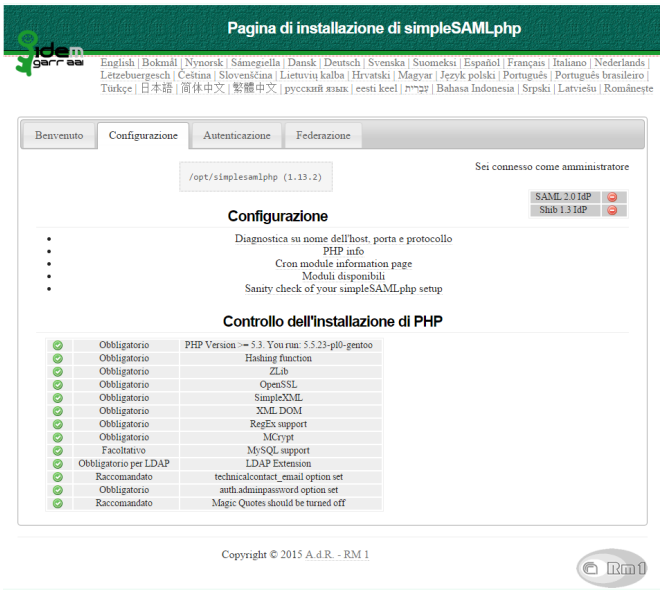


Fig. 3 Interfaccia di controllo SimpleSamlPHP

Una volta conclusa la fase di Test il SP appena creato entrerà a far parte della lista dei Service Provider autorizzati dalla Federazione abilitando così lo stesso allo scambio di asserzioni SAML, con gli IdP già registrati in IDEM, volte all’ Identificazione e Autenticazione degli utenti.

Questo tipo di asset prevede che ogni entità che si registra con la federazione per fornire identità digitali (IDP) fornisca queste tramite un apposito server tipicamente gestito in casa. Questo fa sì che l’utente che vorrà autenticarsi tramite il nostro Captive Portal, per ottenere l’accesso a internet, deve essere messo nella condizione di poter raggiungere l’Identity Provider di appartenenza.

Il Captive Portal, come già spiegato precedentemente, inibisce le connessioni verso qualsiasi host fintanto che l’utente non viene identificato, autenticato e autorizzato. Per poter ottemperare la procedura di identificazione e autenticazione Federata è stata sfruttata la capability chiamata “Walled Garden” messa a

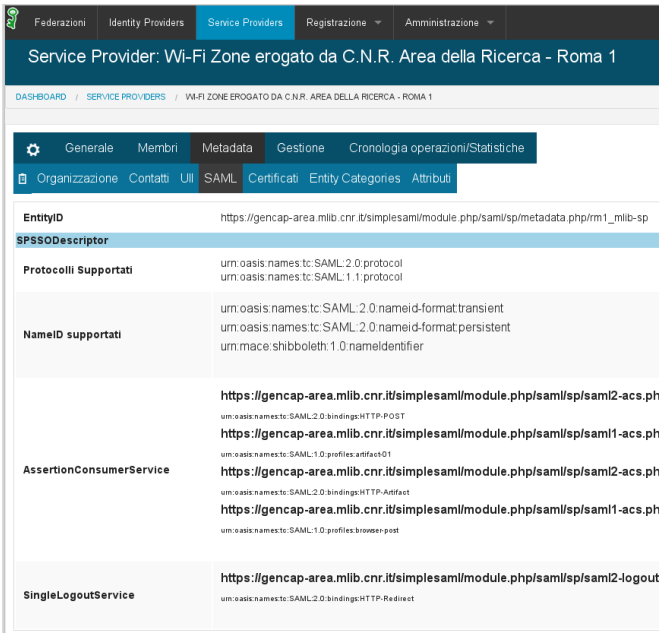


Fig. 4 Portale GARR sezione verifica Metadati

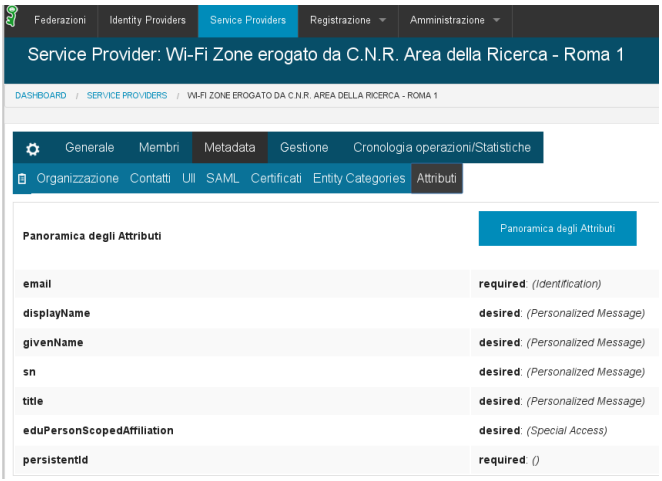


Fig. 5 Portale GARR sezione Attributi richiesti

disposizione da CoovaChilli.

All'interno di CoovaChilli esiste la possibilità di stilare una lista degli host raggiungibili dall'utente a monte dell'autorizzazione altresì obbligatoria (appunto chiamato Walled Garden), la lista degli host concessi può essere implementata per singolo utente tramite gli attributi Radius della connessione oppure può essere stilata all'interno della configurazione di CoovaChilli per essere sfruttata da tutti gli utenti collegati al Captive. Entrambe queste soluzioni sono risultate inefficienti o inutilizzabili al nostro scopo in quanto allo stato attuale si possono contare oltre 3000 Identity Provider (quindi 3000 hosts) e non è pensabile consentire preventivamente la connessione verso un così alto numero di host anche per via del limite stesso del software in uso che prevede un massimo di 1024 slot per lo static garden e altri 1024 per il dynamic garden (max 2048 hosts).

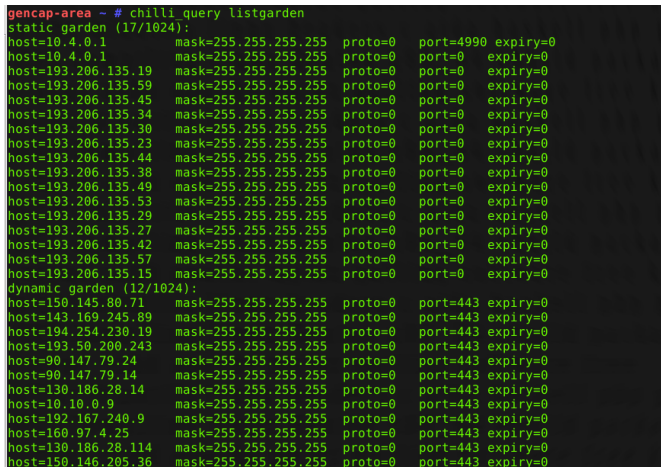
L'altra metodologia non è tuttavia applicabile in quanto è necessario conoscere l'IdP di appartenenza prima che l'utente si colleghi all'HotSpot in modo da comunicare il corretto attributo al server Radius da associare all'utente ancor prima della fase di autenticazione. La soluzione applicata in questo frangente consiste nello sviluppo di una funzione, in codice PHP, che intercetta la scelta dell'IdP di appartenenza da parte dell'utente dalla pagina di WAYF (Where Are You From) del nostro SP e l'aggiunta dell'IdP selezionato al garden globale di CoovaChilli come mostrato in figura 6.

```
<?php

function addgarden($entity){
        $addo=parse_url($entity);
        $garden=$addo['host'];
        if (!isset($addo['port'])){
                $addo['scheme']=="https" ? $port="443" : $port="80";
                }
        else{
                $port=$addo['port'];
        }
        shell_exec("/usr/sbin/chilli_query addgarden data $garden:$port");
}
isset($_GET['idpentityid'])?addgarden($_GET['idpentityid']):null;
$discoHandler = new SimpleSAML_XHTML_IdPDisco(array('saml20-idp-remote', 'shib13-idp-remote'), 'saml');
$discoHandler->handleRequest();
~
~
~
~
```

Fig. 6 Funzione PHP per l'aggiunta dell'IdP al Walled Garden

Questa funzione si avvale del comando "addgarden" della suite CoovaChilli ed è stata inserita all'interno del file disco.php, presente nell'installazione di SimpleSamlPhp, garantendo così l'accessibilità esclusivamente verso l'Identity Provider di turno.



```
gencap-area ~ # chilli_query listgarden
static garden (17/1024):
host=10.4.0.1          mask=255.255.255.255  proto=0   port=4990 expiry=0
host=10.4.0.1          mask=255.255.255.255  proto=0   port=0    expiry=0
host=193.206.135.19    mask=255.255.255.255  proto=0   port=0    expiry=0
host=193.206.135.59    mask=255.255.255.255  proto=0   port=0    expiry=0
host=193.206.135.45    mask=255.255.255.255  proto=0   port=0    expiry=0
host=193.206.135.34    mask=255.255.255.255  proto=0   port=0    expiry=0
host=193.206.135.30    mask=255.255.255.255  proto=0   port=0    expiry=0
host=193.206.135.23    mask=255.255.255.255  proto=0   port=0    expiry=0
host=193.206.135.44    mask=255.255.255.255  proto=0   port=0    expiry=0
host=193.206.135.38    mask=255.255.255.255  proto=0   port=0    expiry=0
host=193.206.135.49    mask=255.255.255.255  proto=0   port=0    expiry=0
host=193.206.135.53    mask=255.255.255.255  proto=0   port=0    expiry=0
host=193.206.135.29    mask=255.255.255.255  proto=0   port=0    expiry=0
host=193.206.135.27    mask=255.255.255.255  proto=0   port=0    expiry=0
host=193.206.135.42    mask=255.255.255.255  proto=0   port=0    expiry=0
host=193.206.135.57    mask=255.255.255.255  proto=0   port=0    expiry=0
host=193.206.135.15    mask=255.255.255.255  proto=0   port=0    expiry=0
dynamic garden (12/1024):
host=150.145.80.71     mask=255.255.255.255  proto=0   port=443 expiry=0
host=143.169.245.89    mask=255.255.255.255  proto=0   port=443 expiry=0
host=194.254.230.19    mask=255.255.255.255  proto=0   port=443 expiry=0
host=193.50.200.243    mask=255.255.255.255  proto=0   port=443 expiry=0
host=90.147.79.24      mask=255.255.255.255  proto=0   port=443 expiry=0
host=90.147.79.14      mask=255.255.255.255  proto=0   port=443 expiry=0
host=130.186.28.14     mask=255.255.255.255  proto=0   port=443 expiry=0
host=10.10.0.9         mask=255.255.255.255  proto=0   port=443 expiry=0
host=192.167.240.9     mask=255.255.255.255  proto=0   port=443 expiry=0
host=160.97.4.25       mask=255.255.255.255  proto=0   port=443 expiry=0
host=130.186.28.114    mask=255.255.255.255  proto=0   port=443 expiry=0
host=150.146.205.36    mask=255.255.255.255  proto=0   port=443 expiry=0
```

Fig. 7 Lista degli host aggiunti al Walled Garden

L'unico limite di questa implementazione è che qualora si presentassero, nella stessa giornata, più di 1024 utenti che richiedono accesso presso altrettanti diversi IdP il 1025 esimo utente non riuscirebbe a contattare il proprio IdP. Poiché questi casi sono statisticamente molto improbabili, riteniamo che la soluzione adottata sia sufficientemente valida.

Al termine dell'autenticazione da parte dell'IdP, un token SAML viene generato e spedito al SP che lo aveva generato, il quale a seguito della verifica del token garantisce o meno l'accesso al servizio; nel nostro caso l'accesso alla rete. Per implementare le dinamiche AAI fornite da SAML con quelle del Captive Portal, a seguito dell'autenticazione da parte dell'IDP si reindirizza il client ad un'apposita landing page che contiene la porzione di codice necessaria alla verifica del token SAML per l'eventuale autorizzazione dell'utente che tenta di accedere.(Fig. 8) I casi d'uso di questo tipo di accesso sono prevedibilmente durante le conferenze che si svolgono presso il Campus o la presenza come ospite all'interno di una determinata struttura, questo presupposto rende evidente che un eventuale utente può sfruttare il medesimo accesso da più dispositivi anche contemporaneamente (ad es. un pc portatile ed il cellulare).

Fig. 8 Codice PHP di verifica token SAML

Per questo motivo è stato di proposito disabilitato lo switch di controllo sessione singola presente nel Radius.

Per dare la possibilità a tutti gli utenti che si collegano di controllare gli accessi: quando e quanti, ma soprattutto da quale dispositivo sono stati effettuati tramite le proprie credenziali è stata realizzata un pagina di riepilogo(Fig. 10), dove è possibile visionare tutti gli accessi effettuati con particolare attenzione alla data, durata della connessione e al MAC-Address del dispositivo utilizzato. Per accedere a questa schermata basta fare click sul link “More info” (indicato dalla freccia rossa nella figura 9) dalla pagina di benvenuto una volta effettuato l’accesso.

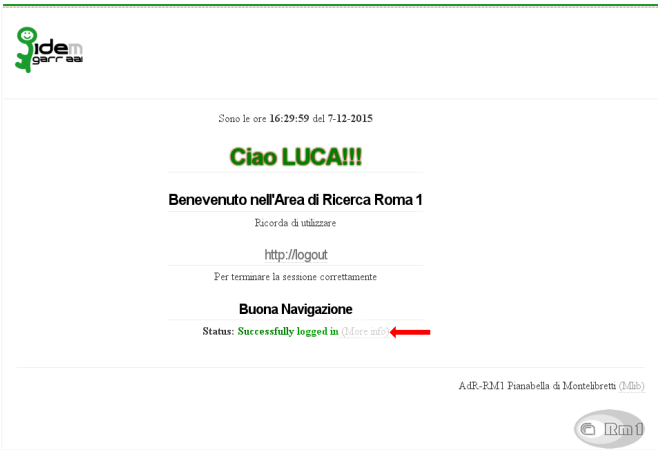


Fig. 9 Messaggio di benvenuto, in evidenza il link alla pagina di riepilogo accessi

La struttura del Portale fa si che tutti gli utenti che ottengono l’accesso alla navigazione vengano identificati su internet attraverso lo stesso indirizzo IP. Per garantire comunque la rintracciabilità dell’utilizzatore finale, nel caso di eventuali infrazioni delle normative vigenti, e stato installato e configurato un software, sempre di matrice Open Source, dal nome Conntrackd.

Questo demone si occupa di registrare l’indirizzo IP sorgente e quello di destinazione di tutte le connessioni che avvengono attraverso il Captive Portal.

Durante il primo accesso alla procedura AAA qui illustrata, è possibile incappare in un errore non gestibile in maniera automatizzata: se l’utente che fa accesso al portale tenta di raggiungere una pagina con connessione protetta (HTTPS) il portale non sarà in grado di effettuare correttamente il redirect alla pagina di login, si otterrà invece un errore di errato certificato e pagina non raggiungibile a causa delle dinamiche SSL e dell’accesso ancora non garantito. In CoovaChilli è previsto il redirect per le connessioni SSL ma anche abilitando questa opzione si incapperebbe in un errore dei certificati di sicurezza, anche per questo motivo è preferibile non abilitare questa opzione. L’utente che vuole effettuare l’accesso dovrà provare ad accedere ad una pagina non protetta da SSL per essere correttamente reindirizzato alla pagina di login.

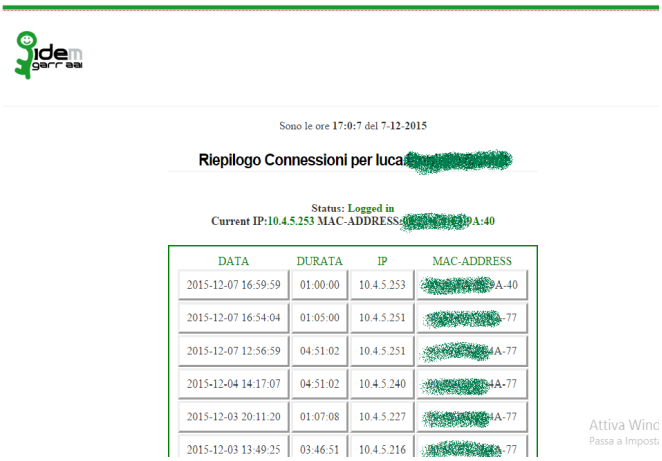


Fig. 10 Pagina riepilogo accessi effettuati

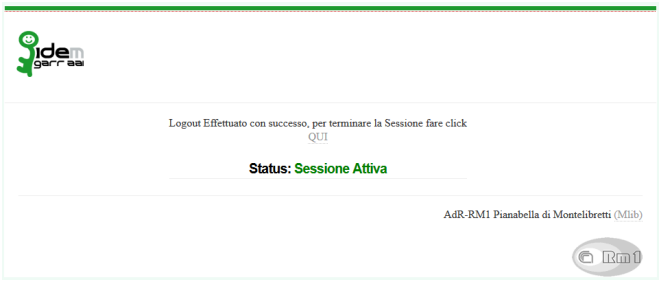


Fig. 11 Pagina di verifica sessione a seguito del logout

Un’altra accortezza che occorre utilizzare nel caso più utenti utilizzino la medesima PDL in condivisione è la procedura di Logout: una volta effettuato l’accesso tramite il portale viene instaurata una sessione che rimane attiva fintanto che l’utente mantiene aperto il browser, questo comporterebbe che un utente potrebbe navigare attraverso il portale tramite la sessione ancora attiva dell’utente che lo ha preceduto. I browser di ultima generazione, come Chrome, hanno sviluppato una capacità intrinseca che permette ad un eventuale utente di mantenere la sessione attiva nonostante il browser venga chiuso e successivamente riaperto, molto utile per l’utente singolo che non dovrà reimmettere le proprie credenziali ove richieste, male invece nel caso di una PDL utilizzata da più utenti, in questo caso viene consigliato l’utilizzo di Mozilla Firefox, che se correttamente utilizzato, evita di lasciare sessioni attive riutilizzabili da altri utenti che potrebbero navigare con la sessione dell’utente che lo ha preceduto. Per chiudere correttamente la sessione con il Captive Portal è previsto l’utilizzo di un URL speciale: http://logout

Visitando questa url speciale verrà visualizzata la procedura di Logout.

L’utilizzo di questo indirizzo fa si che il SP chiuda la sessione attiva dell’utente, tuttavia si riscontra che se l’IdP utilizzato per l’autenticazione è privo del servizio di SingleLogOut, come nella maggior parte dei casi, lo stesso continua a mantenere la ses-

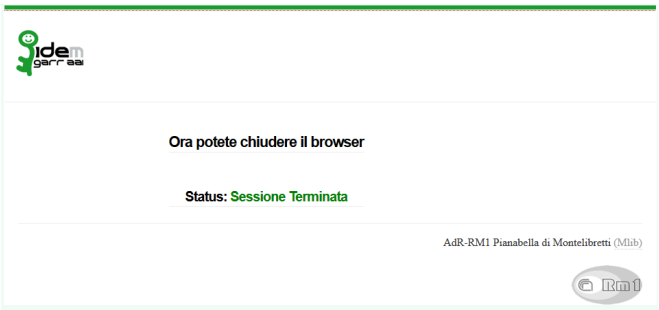


Fig. 12 Messaggio di conferma di avvenuta chiusura sessione

sione utente attiva. In questo caso se dal browser Chrome con il quale è stata effettuato il "logout" e successivamente chiuso, si tentasse di accedere nuovamente tramite il medesimo IdP ed il medesimo Chrome, verrà ripristinata in maniera automatica la sessione dell'utente precedente a causa dell'IdP sul quale la vecchia sessione risulta ancora attiva. Se per la navigazione si utilizza Mozilla Firefox tutto funziona "as aspected".

5 Conclusioni

Anche in questa occasione ciò che risulta evidente è l'estrema flessibilità ed adattabilità dei software utilizzati in completa armonia con l'ambiente che lo contraddistingue: l'OpenSource.

Il servizio di accesso alla rete Federato risulta essere un utile strumento per consentire a coloro che si trovano al di fuori della propria sede operativa di potersi connettere ad internet digitando le proprie credenziali senza dover richiedere permessi ai responsabili delle infrastrutture ospitanti. Soprattutto nell'ambito della ricerca e dell'università, dove esiste una continua e numerosa mobilità, l'accesso alla rete per mezzo di un IdP federato.

Un caso pratico è quello dell'AdR-RM1 ove è presente una Foresteria ed una infrastruttura di accesso wireless, l'Ospite è in grado di collegarsi autonomamente ad Internet utilizzando le credenziali fornite dal proprio Ente di appartenenza qualora questo ultimo appartenga alla Federazione IDEM.

6 Glossario

AAI	(Authentication and Authorization Infrastructure)
AUP	(Acceptable Use Policy)
IDEM	(IDEntity Management per l'accesso federato)
RADIUS	(Remote Authentication Dial-In User Service)
SAML	(Security Assertion Markup Language)
SP	(Service Provider)
IdP	(Identity Provider)
PDL	(Postazione Di Lavoro)

Riferimenti

1 A. Pifferi, G. Nantista, L. Ianniello, C. Ricci, L. Rossi, M. Simonetti, Un captive portal per l'utenticazione su reti wifi dedicate agli internet access point liberi., Smart eLab 2 (2013) 15–19. doi:10.30441/smart-elab.v2i0.55.
2 https://www.idem.garr.it/.
3 https://www.coova.github.io/.
4 http://saml.xml.org/saml-specifications/.
5 https://www.gentoo.org/.
6 http://freeradius.org/.
7 https://registry.idem.garr.it/.

Highlighter

Visual Blast.

Giovanni Mele*[a]

BLAST (Basic Local Alignment Search Tool) is a popular program that retrieves a library of sequences that resemble the query. The major problem experienced by new users of BLAST revolves around constructing syntactically and semantically correct command line, getting input files into acceptable formats and assessing the output. Here, we present Visual BLAST a Graphical User Interface to perform BLAST searches. Visual BLAST aims to make BLAST searches accessible to a wider audience with no bioinformatics skill and to facilitate usage among the existing.

Keywords: Sequence Analysis, Graphic User Interface, Software.

1 Introduction

BLAST remains one of the most widely tools used in computational biology. This popular common line program was developed by Stephen Altschul, Warren Gish, Webb Miller, Eugene Myers, and David J. Lipman[1]. BLAST allows comparing a DNA or protein sequence query with a database of sequences, and consents retrieving a library of sequences that resemble the query. BLAST addresses fundamental problems needed in computational biology research. In fact, BLAST can be used for several purposes. These include identifying species, locating domains, establishing phylogeny, DNA mapping, and comparisons.

The BLAST program can either be free available for download at http://blast.ncbi.nlm.nih.gov/Blast.cgi/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=Download and run as a command-line utility "blastall" or accessed for over the web at http://blast.ncbi.nlm.nih.gov/Blast.cgi. The BLAST web server, hosted by the NCBI, allows anyone with a web browser to perform similarity searches against constantly updated databases of proteins and DNA that include most of the newly sequenced organisms.

In the genomic era, the necessity to search several thousand of sequences in a single query makes the utilization of web-hosted programs not feasible for both low speed over internet and for problems correlated with server traffic and stability. At the same time, the BLAST utility that run on local computers has the main drawback in the complex and long command-line string required to run the analyses. In fact, command-line computing environments are very challenging for users without programming experience. Consequently, there is a pressing need for a menu-driven or Graphical User Interface to allow biologists to access the methodology without becoming programmers. To address these and others issues, Visual BLAST, a Graphical User Interface for BLAST searches was developed.

2 Description

Visual BLAST is a user-friendly GUI written in C# and it was developed to simplify the parameters setting for BLAST searches. Visual BLAST implements BLAST allowing the utilization of blank space in the input and output file name and in the folder path. Moreover, Visual BLAST places in the first line of the output tabular text file the headers for each output option selected by the user. Finally, the GUI of Visual BLAST save the last user input settings allowing easier and more efficient searches when running multiple analyses with the same parameters. The application consists of one window with two tabs: *Blast Search and Database Management* (Figure 1).

2.1 Blast Tab

The *Blast Search* tab consists of four panels (Fig. 1A). The Query File panel contains a drag and drop box that accepts the input file for the search in FASTA format. On the right, the *Parameters* panel consists of four dropdown boxes and a checklist box. The dropdown boxes allow setting the number of hits for each input sequence, the type of search (BLASTn for nucleotide query on nucleotide database, BLASTp for aminoacid query on protein database and BLASTx for nucleotide query on protein database), the output file (Table 1) and the minimum e-Value cutoff. In case that the user selects as output the file format option 6, 7 or 10, the checklist box allows to select a custom format specifiers (Table 2). The default specifiers are: *qseqid, sseqid, pident length mismatch gapopen qstart qend sstart send*

[a] *Institute of Agricultural Biology and Biotechnology, National Council of Research, Via Salaria Km. 29.300 00015 Monterotondo Scalo, Rome, Italy.*
* **Contact:** melegio@ibba.cnr.it.

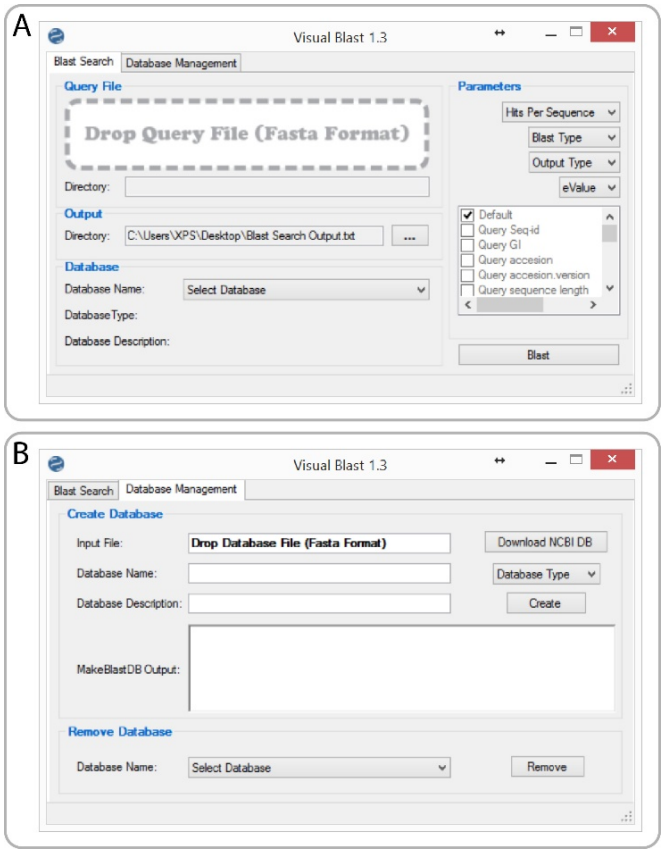


Fig. 1 Visual BLAST Tabs. A) Blast Search Tab. B) Database Management Tab.

evalue. The Output panel allows selecting the output directory and output file name. The *Database* panel allows the selection of the database for the query.

Tabella 1 Output file format.

Selection	Output file format description
0	Pairwise
1	Query-anchored showing identities
2	Query-anchored no identities
3	Flat query-anchored, show identities
4	Flat query-anchored, no identities
5	XML Blast output
6	Tabular with headers
7	Tabular with comment lines
8	Text ASN.1
9	Binary ASN.1
10	Comma-separated values
11	BLAST archive format (ASN.1)

2.2 Database Tab

The *Database Management* tab allows the databases administration and consists of an upper panel for the creation of a new database and a lower part that allows the removal of unwanted databases (Figure. 1B). The database creation section comprises of a drag and drop box for the creation of the new database, of two text boxes for the database information and of a drop down box for the selection of database type (nucleotide or proteic). Moreover, the *Download NCBI DB* button open the NCBI FTP page for the downloading of the NCBI curated prebuild databases. The lower panel of the tab allows the selection and the deletion of unwanted database.

Tabella 2 Custom format specifiers.

Specifiers	Description
qseqid	Query Seq-id
qgi	Query GI
qacc	Query accesion
qaccver	Query accesion.version
sseqid	Subject Seq-id
sallseqid	All subject Seq-id(s), separated by a ';'
sgi	Subject GI
sallgi	All subject GIs
sacc	Subject accession
saccver	Subject accession.version
sallacc	All subject accessions
qstart	Start of alignment in query
qend	End of alignment in query
sstart	Start of alignment in subject
send	End of alignment in subject
qseq	Aligned part of query sequence
sseq	Aligned part of subject sequence
evalue	Expect value
bitscore	Bit score
score	Raw score
length	Alignment length
pident	Percentage of identical matches
nident	Number of identical matches
mismatch	Number of mismatches
positive	Number of positive-scoring matches
gapopen	Number of gap openings
gaps	Total number of gaps
ppos	Percentage of positive-scoring matches
frames	Query and subject frames separated by a '/'
qframe	Query frame
sframe	Subject frame
btop	Blast traceback operations (BTOP)
qseqid	Query Seq-id

3 Conclusion

Visual BLAST executable is free of charge software that runs on 32 and 64 bits Windows platform and it is tune up for Windows XP, 7 and 8. It is distributed in pre-packaged self-extracting installer for installation on local computers. Visual BLAST is a user friendly and flexible GUI software to retrieve a library of sequences that resemble the query. The easy to use GUI Interface benefits a wide audience for a fast and effective analysis.

Availability and Implementation: Visual BLAST executable is freely available on the web page of the Institute of Agricultural Biology and Biotechnology of National Council of Research (http://www.ibba.mlib.cnr.it/Visual_Blast.html), Softpedia (http://www.softpedia.com/get/Science-CAD/Visual-Blast.shtml) and Softtonic (http://visual-blast.en.softonic.com). This software is design to be fully compatible with Windows XP, 7, 8 and 10 environments.

4 Acknowledgements

Special thanks to Donato Giannino e Giulio Testone for the critical reading of the paper. This work was supported by a dedicated grant from the Italian Ministry of Economy and Finance to the National Research Council for the project "Innovazione e Sviluppo del Mezzogiorno - Conoscenze Integrate per Sostenibilità ed Innovazione del Made in Italy Agroalimentare - Legge n. 191/2009"

Riferimenti bibliografici

1 S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool, Journal of Molecular Biology 215 (3) (1990) 403 – 410. doi:10.1016/S0022-2836(05)80360-2.

E-Learning & Scientific Comunication

In materia di diritto d'autore oggi.

Gisella Menichelli,[*a] Antonella Cecchetti,[b] Elisabetta Ceccarelli[c]

Sulla tematica del diritto d'autore ed il mondo delle biblioteche c'è molto da sapere. Nel corso del tempo ci sono stati affinamenti legislativi per la tutela delle opere e la loro fruizione. Le biblioteche che hanno un ruolo importante nella circolazione delle idee, ne devono tener conto. Per questo abbiamo ritenuto interessante farne una panoramica di aggiornamento.

Keywords: Diritto d'autore, copyright, copyleft, biblioteche.

1 Introduzione

C'è molto da sapere ed imparare in termini giuridici e di comportamento sulla materia del diritto d'autore ai nostri giorni. Lo era stato in passato quando il materiale da tutelare era in supporti cartacei, dischi, tele o altro, ma a maggior ragione ora che viviamo in un mondo digitale e online.

Prima di tutto esaminiamo il concetto di copyright e diritto d'autore: il primo è la tutela della proprietà intellettuale nei paesi anglosassoni (Stati Uniti e GB), il secondo è il diritto giuridico di settore in Italia. Non si equivalgono perché ogni paese ha le sue differenze, anche se in linea di principio sono assimilabili.

Ma un conto è l'idea, un conto è la rappresentazione dell'idea. Quello che viene tutelato con il **diritto d'autore è la forma espressiva dell'idea, non il suo concetto**.

2 Discussione

La forma espressiva di un'idea si compone di due voci: **il prodotto** ed **il suo supporto** che in termini di diritto d'autore tutelano i diritti morali e i diritti patrimoniali.

I **diritti morali** sono per loro natura imprescrittibili, irrinunciabili, inalienabili e quindi illimitati nel tempo: durano per sempre e si ereditano; ma non sono fonte di guadagno.

L'opera d'ingegno per essere tutelata nel sistema legislativo del diritto d'autore italiano deve essere:

1. originale
2. non necessita di registrazione alcuna. Se però s'intende tutelare un'opera dal ricorso al probatorio, cioè dover dimostrare con prova l'autenticità della sua paternità - in questo caso, si può ricorrere a: spedizione postale con timbro e sigillo, spedizione PEC (Posta Elettronica Certificata), deposito notarile, SIAE (Società Italiana Autori e Editori)[1] e altro. Da tener presente che il deposito delle opere inedite ha finalità probatorie e non costitutive del diritto d'autore. Il deposito può avvenire anche da parte di soggetti non iscritti alla Società Italiana Autori Editori.
3. è estensibile agli editori e produttori attraverso i diritti connessi.

Nel **sistema internazionale del copyright** invece l'opera:

- deve essere registrata ©, ma senza più obbligo dal 1998 (DMCI)
- può avere un'originalità limitata
- si offrono maggiori tutele al diritto economico piuttosto che a quello morale.

L'autore può cedere in parte o completamente lo sfruttamento dei diritti economici di una sua opera ad una terza figura che può essere l'editore, dietro compenso o no. I **diritti economici o patrimoniali** sono dei diritti esclusivi che solo l'autore può scegliere di cedere o autorizzare a farlo. Quindi possono essere venduti, acquistati, regalati oppure trasmessi in asse ereditaria. Che cosa può fare chi detiene i **diritti esclusivi patrimoniali** di un'opera?

- può pubblicarla,
- può sfruttarla economicamente,
- può riprodurla,

[a] *CNR-Istituto dei Sistemi Complessi, Biblioteca di Area, Area della Ricerca di Roma 1, Via Salaria Km 29,300, Monterotondo Scalo (Roma).*
[b] *CNR-Istituto per lo Studio dei Materiali Nanostrutturati, Biblioteca di Area, Area della Ricerca di Roma 1, Via Salaria Km 29,300, Monterotondo Scalo (Roma).*
[c] *CNR-Istituto di Struttura della Materia, Biblioteca di Area, Area della Ricerca di Roma 1, Via Salaria Km 29,300, Monterotondo Scalo (Roma).*

- può trascriverla,
- può eseguirla, rappresentarla o recitarla,
- può diffonderla,
- può comunicarla al pubblico,
- può distribuirla e metterla in commercio,
- può tradurla,
- può inserirla in una raccolta,
- può rielaborarla,
- può noleggiarla o prestarla.

Tutti questi diritti sono autonomi e quindi per ognuno è richiesta una specifica autorizzazione all'autore, se si vuole procedere ad es. con una traduzione e con una rielaborazione: sono due le autorizzazioni distinte da richiedere.

Abbiamo detto già che il diritto patrimoniale dura tutta la vita e 70 anni dopo la morte dell'autore.

La legge del diritto d'autore, in Italia, fu scritta il 22 Aprile 1941 con il n. 633: *Protezione del diritto d'autore e di altri diritti connessi al suo esercizio*, poi nel corso del tempo ha subito delle modifiche e aggiustamenti per adeguarsi al mondo in evoluzione. Abbiamo le modifiche inerenti alla diffusione telematica abusiva con la legge del 22 maggio 2004, n. 128, quelle inerenti alle università e alla ricerca con la legge 31 marzo 2005, n. 43, e ancora quelle inerenti alla libera pubblicazione in rete di opere musicali o artistiche a bassa risoluzione o degradate con il DDL S861 del 21 febbraio 2007.

Cosa tutela questa legge?

1. I diritti morali e perenni della proprietà intellettuale, della paternità dell'opera che viene anche ereditata;
2. I diritti patrimoniali, ovvero di sfruttamento economico che durano tutta la vita e 70 anni dopo la morte dell'autore.

Tra i diritti morali vi sono: l'integrità quindi il diritto di non modificare l'opera per come è stata concepita e la pubblicazione come opera prima, l'inedito ma anche la traduzione.

Tra i diritti patrimoniali vi sono: la riproduzione, la trascrizione, la recitazione, la distribuzione, la traduzione, la pubblicazione in una raccolta, la rielaborazione, il prestito, ecc.

Ma oltre al diritto d'autore vi sono opere che ricadono nel **pubblico dominio**, per decorrenza dei termini di tutela previsti dall'attuale legge: 70 anni dalla morte dell'autore o dell'ultimo autore se si tratta di un opera collettiva (es. un film) oppure perché sono **documenti di fonte pubblica** (ad es. le Leggi), oppure semplicemente per opere non creative.

Ad es. nel pubblico dominio nel 2015 sono entrate le opere di autori come:

Antoine de Saint.Exupéry, Vasilij Kandinsky, Edvard Munch; Filippo Tommaso Marinetti, ecc. Il simbolo che viene usato per indicare che un'opera è nel pubblico dominio è questo: , ma attenzione! Il contenuto soltanto è nel pubblico dominio, le illustrazioni potrebbero non esserlo.

Infatti, le immagini sul web per essere utilizzate devono essere di pubblico dominio, altrimenti bisogna chiederne autorizzazione al detentore dei diritti. In Google Immagini si può selezionare la ricerca in base alla licenza d'uso.

Il concetto di **Fair use** ovvero delle **libere utilizzazioni** è tipico del sistema legislativo del Copyright, detto anche **Fair dealing**, non propriamente del diritto d'autore italiano e prevede appunto le **libere utilizzazioni** oppure le **limitazioni dei diritti**. Sono **eccezioni alle regole** di sfruttamento economico dell'opera per poterla riusare in parte, rielaborandola in opere creative o per motivi di studio o di ricerca. (Art. 70, legge 633, 1941 e successive modifiche). Quindi è consentito farne riassunti, citazioni, o comunicazioni in pubblico purché sia a fini critici o di discussione.

Il Fair use si basa su quattro fattori:

- non deve essere a scopo di lucro,
- deve mantenere la natura dell'opera protetta,
- dipende da quanta parte dell'opera viene usata,
- dipende da quanto può incidere economicamente sull'opera protetta.

Dal punto di vista delle **armonizzazioni legislative** tra i vari paesi che compongono il panorama europeo e internazionale abbiamo diverse azioni per ogni diverso paese. Un primordiale documento importante a livello internazionale è la **Convenzione di Berna** sulla protezione delle opere letterarie e artistiche del 1886: riconosce il diritto d'autore tra le nazioni convenzionate.

Al pari la precedente **Convenzione di Parigi** (1883) si rifà alla proprietà industriale con marchi e brevetti. L'Italia partecipa alla Convenzione di Berna dal 1961.

All'interno dell'Unione Europea vengono redatte otto direttive sul **Copyright and Neighbouring Rights**, che regolamentano la proprietà intellettuale non brevettuale:

1. il prestito e il noleggio,
2. la durata dei diritti,
3. le banche dati,
4. i programmi per elaboratori,
5. la comunicazione al pubblico,
6. internet,
7. le opere d'arte,
8. il rafforzamento delle tutele.

Interessante è la direttiva europea nota come **Ipred2** del 2007. Questa introduce l'emendamento in cui *la riproduzione in copie o su supporto audio o qualsiasi altro mezzo, a fini di critica, recensione, informazione, insegnamento (compresa la produzione di copie multiple per l'uso in classe), studio o ricerca, " non debba essere qualificato come reato"*.[2]

Per quanto attiene le **biblioteche** le eccezioni al diritto d'autore nelle modifiche apportate alla legge 633 del 1941, relative alle **Utilizzazioni libere e nella Sezione I - Reprografia**

... consentono: il riassunto, la citazione, la riproduzione o comunicazione di brani a scopo didattico o di studio e ricerca.

E' possibile quindi per una biblioteca ed i suoi utenti utilizzare il materiale informativo (capitoli, articoli, estratti, ecc.) protetto dal diritto d'autore quando l'uso non è a scopo di lucro, ma è di studio o di ricerca, per il bene collettivo.

In particolare l'Art. 68, innovato da vari correttivi legislativi (D.Lg. 9 aprile 2003, n. 68), prevede:

- una restrizione alle attività reprografiche come da modifica della legge n. 248 del 2000,
- il recepimento della Sesta Direttiva europea che armonizza l'analogico (il supporto cartaceo) con il digitale.

Questa Direttiva che tutela il diritto di riproduzione incluse le riproduzioni digitali in qualsiasi forma e supporto, ha ovviamente limitato fortemente la possibilità di accesso al documento in ambito bibliotecario.

L'Art. 68 si limita alla riproduzione della sola copia cartacea e si rifà alla **"limitazione dei diritti" dietro corresponsione di compenso** anche senza autorizzazione. Riassumendo le limitazioni ai diritti degli autori o degli editori se vi è stata cessione in parte o totale e senza doverne chiedere autorizzazione o senza doverne corrispondere un compenso sono:

1. la riproduzione di articoli di carattere economico, politico, religioso o di attualità pubblicati su rivista (basta indicarne la fonte);
2. la riproduzione di discorsi politici o amministrativi (basta indicarne l'autore, il luogo e la data);
3. la riproduzione di opere in procedure giudiziarie (basta indicarne la fonte);
4. la citazione, il riassunto o la riproduzione di brani o parti di opere non a fini di lucro (basta citarne la fonte).

Per quanto riguarda la **reprografia** (riproduzione meccanica in fotocopia o altro simile) questa si effettua senza autorizzazione, ma con compenso (in Italia su base forfettaria alla SIAE). Infatti, l'art. 68 differenzia le attività di riproduzione per i servizi di biblioteca da quelle per uso personale.

Tra le varie stesure e le ambiguità conseguenti alle interpretazioni dell'Art. 68, si possono considerare come **Linee guida** questi punti:

- le fotocopie si possono effettuare unicamente a scopo di studio o ricerca, per uso personale e fino al 15% del volume o fascicolo totale, esclusa la pubblicità.
- L'utente si assume ogni responsabilità per l'uso che ne farà, essendo severamente vietata qualsiasi successiva riproduzione o pubblicazione ad uso commerciale.

La biblioteca del CNR non è un Centro Copia, bensì un servizio alla ricerca, per cui le fotocopie ai ricercatori o altro personale interno fanno parte del servizio stesso e quindi si applica la seguente norma:

"E' libera la fotocopia da opere esistenti nelle biblioteche, fatta per i servizi della biblioteca o, nei limiti e con le modalità di cui ai commi quarto [riguarda i Centri Copia e l'equo compenso alla SIAE] e quinto [riguarda le biblioteche e il servizio per uso didattico], per uso personale".

La riproduzione di interi volumi è severamente vietata, salvo eccezioni di rarità, fuori catalogo, presenti in biblioteca, ma che al loro volta vista la rarità e per garantirne l'incolumità ne potrebbe essere vietata la riproduzione stessa. Il tipico caso del "lo puoi fare, ma non lo puoi fare".

Adesso veniamo alla **copia temporanea in digitale**. Viene aggiunta alla legge sul diritto d'autore (9 aprile 2003,n.68) con l'Art. 68 bis, dopo il recepimento della Direttiva europea 2001/29 CE, secondo cui non costituiscono violazione del diritto di riproduzione *"gli atti di riproduzione temporanea privi di rilievo economico proprio che sono transitori o accessori e parte integrante e essenziale di un procedimento tecnologico, eseguiti all'unico scopo di consentire la trasmissione in rete tra terzi con l'intervento di un intermediario, o un utilizzo legittimo di un'opera o di altri materiali"*.

Qualche nota sul servizio **NILDE**[3] (Network Inter Library Document Exchange) del Cnr di Bologna, divenuto con il tempo sistema di cooperazione tra biblioteche per il document delivery.

All'interno del servizio vi è **ALPE** (Archivio Licenze Periodici Elettronici), database sui contratti di licenza. Si possono cercare informazioni sull'uso consentito dalle licenze attraverso una maschera di ricerca per ISSN (International Standard Serial Number) o ISBN (International Standard Book Number) oppure attraverso l'Editore. Selezionando la licenza si visualizzeranno le informazioni semplificate delle clausole al DD/ILL.

Ad es.: per un editore possono valere queste regole

- **DD/ILL consentito**: Sì
- **Metodi di invio**: posta, fax o invio elettronico
- **Formato del documento da inviare**: File originale dell'editore
- **Indicazioni per biblioteca richiedente**:
  1. Obbligo di cancellare il file subito dopo la stampa: ✓
  2. La richiesta dell'utente è esclusivamente per scopi di ricerca o di studio personale: ✓
  3. Formato documento per l'utente finale: copia cartacea.
- **Indicazioni per biblioteca fornitrice**: ✓
  1. Il servizio DD/ILL non può essere a fini commerciali.

E' responsabilità dell'operatore di biblioteca rispettare gli usi consentiti dalle licenze nella pratica di DD/ILL.

Inoltre, ci sono le Linee Guida CONTU prodotte dalla National Commission on New Technology Uses of Copyrighted Works del 1978, nelle quali sono imposti dei limiti al numero di copie che possono essere trasmesse da una biblioteca. Stabiliscono che il servizio DD non deve essere sostitutivo dell'acquisto. Ovvero non sono ammesse più di cinque copie l'anno dello stesso articolo di periodico pubblicato negli ultimi cinque anni dalla data della richiesta e per la stessa istituzione.

Le copie trasmesse in sistemi sicuri (SEDD - Secure Electronic Documet Delivery) sono ammesse nelle licenze d'uso nella forma dei "sistemi analoghi" di trasmissione. NILDE pur non essendo citata direttamente tra i "sistemi analoghi" delle trasmissioni sicure, di fatto lo è.

La sottoscrizione dei contratti in licenza d'uso per le risorse elettroniche avviene attraverso l'azione di coordinamento **CRUI/CARE**[4] **e CIPE**[5]. CARE - Coordinamento per l'Accesso alle Risorse Elettroniche della CRUI – Conferenza dei rettori delle Università Italiane http://www.crui-risorselettroniche.it/

La Commissione Biblioteche CRUI e i Consorzi CASPUR, CIBER, CILEA, CDL e CIPE nel 2005 ha sottoscritto una Convenzione per *"favorire il raggiungimento di economie nell'acquisto e nella gestione delle risorse elettroniche"* e nel corso del tempo al gruppo si sono aggiunti sempre più esperti di settore.

Altri compiti del CARE sono a livello negoziale di ottenere i miglior benefici all'accesso informativo e avere un rafforzamento contrattuale con le potenti lobbies editoriali.

Tra gli utenti CARE rientra il CNR.

Il CIPE è Il Consorzio Inter-istituzionale per Progetti Elettronici – Bibliotecari, Informativi e Documentari di alcune Università in particolare che si sono riunite sotto la stessa convenzione: riguardano le Università di Ancona, Bologna, Modena, Reggio Emilia, Parma, Firenze, Pisa, Siena, Genova, Padova e Venezia Ca' Foscari.

Le biblioteche possono **prestare le opere** coperte dal diritto d'autore in base all'art. 9 della legge 633 del 1941. Quest'articolo prevede il libero prestito con alcune eccezioni (come gli spartiti, oppure le opere cinematografiche se ancora non distribuite, ecc.). La Commissione europea prevede invece di far remunerare gli autori anche nel caso del prestito. In Italia le biblioteche afferenti al MIUR (quindi Scuola, Università e Ricerca) ne sono esenti. Per le altre biblioteche, dovrebbe essere il Ministero per i Beni e le Attività Culturali e con il concorso delle Regioni a sostenere il diritto al prestito pubblico nelle biblioteche del territorio.

Novità del 2014: la procedura amministrativa del **Regolamento Agcom n. 680/13/CONS**.

L'Autorità per le Garanzie nelle Comunicazioni ha approvato nel 2014 il regolamento sulla tutela del diritto d'autore online che consente a chi ritiene di aver subito una violazione sulla paternità di un contenuto possa farne denuncia e richiesta immediata di rimozione all'Agcom direttamente e senza passare per le vie legali. Ma se la violazione ha già in essere un giudizio civile questa contestazione rapida non è possibile.[6]

Alcune informazioni aggiuntive sulla **condivisione dei contenuti** in internet:

Il tasto **condividi** su internet non significa di per sé **libero dominio**, ma un'ordinanza della Corte Europea sull'**embedded o incorporamento** di un video o di altro contenuto dice che *"L'incorporamento in un sito web di un'opera protetta, che è pubblicamente accessibile su un altro sito, per mezzo di un collegamento tramite frame di per sé non costituisce una comunicazione al pubblico ai sensi della direttiva Ue sul copyright nella misura in cui l'opera in questione non è comunicata ad un nuovo pubblico, né si utilizza uno specifico mezzo tecnico diverso da quello utilizzato per la comunicazione originale"*, quindi in sostanza non costituisce violazione del diritto d'autore, essendo già pubblico e non renumerato.[7]

Alcune clausole contrattuali dei canali più noti in internet prevedono:

1. L'autore che pubblica dei Contenuti su **YouTube** concede "una licenza per il mondo intero, non esclusiva, gratuita, trasferibile (con diritto a concedere sub-licenze) ad usare, riprodurre, distribuire, preparare opere derivate, visualizzare ed eseguire tali Contenuti" attraverso il Servizio YouTube Player e accetta i termini semplicemente utilizzandolo. Pertanto, **"l'embedding" appare lecito sotto il profilo del diritto di autore**, fatti sempre salvi altri illeciti che si potrebbero configurare (es. concorrenza sleale/violazione della privacy, danni, sfruttamento economico senza autorizzazione, etc.).

2. L'autore che pubblica su **Twitter** concede a Twitter "una licenza che autorizza a rendere i tweet dell'utente disponibili al resto del mondo e autorizza altri soggetti a fare altrettanto."

Adesso parliamo di **copyleft**: il contrario del copyright ovvero una cessione in parte dei diritti alla libera utilizzazione secondo regole stabile dall'autore, ad es. le **Creative Commons**, oppure le **licenze GNU** per il software.

Creative Commons è (Organizzazione no-profit, sovvenzionata da donazioni) che rilascia licenze d'uso e quindi protegge un'opera.

In Italia il sito è: http://www.creativecommons.it/ a cui - a partire dal 2003 - ha contribuito anche il CNR con l'Istituto di Elettronica e di Ingegneria dell'Informazione e delle Telecomunicazioni (IEIIT).

Esistono sei diverse licenze frutto di combinazioni dei diritti che l'autore può decidere di concedere in autorizzazione, quindi ad esempio possono essere esclusi i diritti commerciali (NC), oppure di farne un uso per opere derivate (ND) , o se si autorizzano di concederne l'uso medesimo (SA).

Le licenze Creative Commons sono liberamente e gratuitamente utilizzabili da chiunque.

Con l'attribuzione del codice di licenza **CC Creative Commons 3.0** [oppure n. di versione, anche internazionale] (**BY**) le opere dell'ingegno si possono condividere e modificare in totale libertà anche a fini commerciali a patto che si citi l'autore e si indichi cosa si è modificato. Fermo restando di verificare eccezioni o restrizioni aggiunte.

Della **CC Creative Commons 3.0** abbiamo poi le seguenti specifiche (i **Commons Deed** – note semplificate del contratto (**Legal Code**) che non hanno valore legale, ma servono solo a spiegarne il contenuto):

1. **Non opere derivate 3.0** ovvero non si concede l'uso di farne opere derivate e pubblicarle (**ND**)
2. **Non commerciale 3.0** ovvero non puoi farne un uso commerciale (**NC**)
3. **Non commerciale 3.0 e Non opere derivate 3.0**
4. **Condividi allo stesso modo 3.0** devi redistribuire con la stessa licenza (**SA**) 5. **Non commerciale - Condividi allo stesso modo 3.0**

La **licenza CC0** attribuisce il pubblico dominio, esiste la possibilità di aggiungere qualche ulteriore accordo con la licenza CC Plus.

Il tutto corredato di **metadati** per essere facilmente individuati dai motori di ricerca abilitati:

http://search.creativecommons.org/.

Inoltre è bene sapere che non è necessario registrarsi, basta utilizzare il logo Creative Commons con la dicitura dei diritti espressi e indicati tra quelli delle specifiche:

http://creativecommons.org/choose/

Esempi qui:

https://wiki.creativecommons.org/wiki/Marking_your_work_with_a_CC_license

I fondatori dell'Organizzazione Creative Commons sono stati, nel 2001: James Boyle, Michael Carroll e Lawrence Lessing: esperti di diritto della rete e proprietà intellettuale, Hal Abelson informatico del MIT, Eric Saltzman avvocato esperto della rete ed Eric Eldred editore web per il pubblico dominio e molti altri ancora tra studenti e ricercatori hanno contribuito al suo sviluppo. Da ricordare, tra questi, il contributo al codice di Aaron Swartz.

Infine, in questa pagina possiamo scaricare tutti i loghi rispettando le policies dell'organizzazione:

https://creativecommons.org/about/downloads/

Interessante anche il logo Free Cultural Work che stabilisce una licenza a minor restrizioni sulla base del CC-BY or BY-SA o anche CC0 oppure Public Domain.

Fig. 1

Tutte le altre licenze CC non sono considerate Free Cultural Work perché restringono il campo delle libertà, magari limitandone lo sfruttamento commerciale oppure vietandone la derivazione, ecc. Il logo in questo caso è ❄.

Altra organizzazione che si occupa di armonizzare la legislazione sulla proprietà intellettuale nel mondo si chiama **WIPO: World Intellectual Property Organization** con sede a Ginevra, istituita nel 1967.

Nel 1994 a livello internazionale si è raggiunto un accordo con la **World Trade Organization - WTO** in cui i software e le banche dati furono considerati ascrivibili alle opere dell'ingegno da tutelare secondo la Convenzione di Berna, quindi non come brevetti.

I trattati internazionali vengono recepiti automaticamente in Italia, in quanto paese aderente alla Convenzione di Berna.

In ambito europeo l'armonizzazione sul diritto d'autore ha portato a una serie di direttive dedicate ai nuovi mezzi di comunicazione, già recepite in Italia: sulla durata della protezione, sul diritto di prestito e noleggio, sulla radiodiffusione, sul software e sulle banche dati. Sulle banche dati c'è da dire che possono avere due tipi di tutela, quella del diritto d'autore quando sono creative, e quella del **diritto sui generis** (diritto connesso) quando non lo sono. La durata varia da 70 anni dalla morte, oppure nel secondo caso a 15 anni. La titolarità è dell'autore nel primo caso e del costitutore nel secondo.[8]

La SIAE rilascia anche licenze su opere multimediali e audiovisive. Per rafforzare la tutela è stato creato il Registro pubblico generale delle opere protette presso il DGBIC – Direzione Generale Biblioteche e Istituti Culturali del MiBACT, il cui scopo è probatorio.[9] (vedi istruzioni per il deposito[10]).

Inoltre, tra le eccezioni e limitazioni al diritto d'autore è bene sapere che vi sono alcuni diritti che consentono (a determinati professionisti) delle libertà come il diritto di cronaca, di parodia, di citazione.

Qualche nota sul **contratto di edizione.**

Il diritto editoriale è diverso dal diritto d'autore, è vincolato da un contratto ed ha una durata in relazione al contratto stesso.

E' un contratto con il quale l'autore cede in parte oppure completamente i diritti patrimoniali all'editore che lo pubblica dietro accordo o meno di corresponsione, conservandone però a pieno titolo i diritti morali. (Art. 118 e seguenti della LDA). Ma c'è un'alternativa . . .

Pubblicare in **Open Access** è una scelta che garantisce diffusione, notorietà e successo attraverso le citazioni, avanzamento di carriera e finanziamenti (è prerogativa del programma UE Horizon 2020), ma soprattutto preponderante è la disseminazione di conoscenza nel mondo.

Esistono varie tipologie di pubblicazioni Open Access:

- le *Gold Open Access* – pubblicazioni OA dietro pagamento dell'Article Processing Charge da parte degli autori o Istituzione per la pubblicazione all'editore, anche definite come OpenChoice, ecc.;
- le *Green Open Access* - pubblicazioni completamente OA licenziate CC o secondo le politiche editoriali attraverso l'auto-archiviazione, ecc. verificabili sul sito:

  http://www.sherpa.ac.uk/romeo/search.php?la=en&fIDnum=all&mode=advanced ;
- le *Ibrido Open Access* – ad articoli free OA limitati;

A questo proposito segnaliamo l'esistenza della *Directory of Open Access Journal* – **DOAJ** che raccoglie le pubblicazioni in open access garantite da un controllo di qualità, attraverso l'ISSN, peer-review o editorial board, ecc. https://doaj.org/

Inoltre, è bene ricordare che negli ultimi anni si sono diffusi i cosiddetti **predatory publishers**, secondo la definizione di un bibliotecario di Denver nel Colorado, Jeffrey Beall. Editori che per strategie di mercato spesso usano sistemi fraudolenti per pubblicare contenuti a pagamento.

Non ultima questa nota relativa alla IODL Italian Open Data License 2.0 (I dati aperti per la pubblica amministrazione, sviluppata da FORMEZ PA): una licenza di condivisione, modifica e riuso libero delle informazioni a condizione di indicarne la fonte ed il licenziante e garantendone un uso non ufficiale e soprattutto non ingannevole.

Fonte:

http://www.dati.gov.it/iodl/2.0/

https://it.wikipedia.org/wiki/Italian_Open_Data_License

3 Conclusioni

Per finire, nell'era digitale e globale dove le informazioni sono fruibili solo a chi possiede gli strumenti per accedervi in formato elettronico (vedi gli abbonamenti editoriali, ma anche il PC, la rete, ecc.), è necessario che sia il bibliotecario ad assicurarne l'accesso a tutti e svolgere così appieno il suo ruolo di supporto e ausilio alla ricerca, come custode e diffusore della conoscenza nel rispetto degli usi consentiti.

4 Appendice

Per approfondire la tematica proposta dall'articolo, gli autori suggeriscono la copiosa letteratura in materia di Antonella De

Robbio, esperta bibliotecaria del settore dell'Università di Padova e fonte di riferimento per il nostro lavoro, sul sito Antonella de Robbio, "Diritto d'autore e copyright" e altri siti utili, quali:

- https://it.wikipedia.org/wiki/Creative_Commons
- Manuale di sopravvivenza per musicisti, Sveva Antonini, Josep Coll Rodriguez, (2012) Paolo Emilio Persiani editore – 3^a edizione.
- Copyright e uso delle immagini in rete. Intervista a Edoardo Tedeschi http://pinterestitaly.com/copyright-utilizzo-immagini-in-rete/
- http://biblioteca.bo.cnr.it/index.php/it/formazione/formazione-bibliotecari

Riferimenti

1 https://www.siae.it/autori-ed-editori/i-registri/deposito-opere-inedite.
2 http://it.cyclopaedia.net/wiki/IPRED2.
3 https://nilde.bo.cnr.it/.
4 http://www.crui-risorselettroniche.it/.
5 http://www.unicipe.it/.
6 http://www.agcom.it/tutela-del-diritto-d-autore.
7 http://www.ilfattoquotidiano.it/2014/10/29/ue-corte-di-giustizia-embeddare-i-video-non-e-reato/1177785/.
8 http://www.altalex.com/documents/news/2011/02/17/disposizioni-sui-diritti-del-costitutore-di-una-banca-di-dati-diritti-e-obblighi-dell-utente.
9 Registri di Pubblicità, Formez PA.
10 Barbara Limonta e Giulia Scacco, Guida al deposito e alla registrazione delle opere nel Registro Pubblico Generale delle Opere Protette.

E-Learning & Scientific Comunication

Scrittura Collaborativa Accademica: metodiche e applicazioni tecnologiche.†

Guido Righini,*[a] Augusto Pifferi,[b] Andrea Lora[b]

Il progresso tecnologico compiuto dagli strumenti di comunicazione di Internet ha reso ora possibile realizzare piattaforme informatiche per la scrittura collaborativa. Per i ricercatori disporre di questo strumento è molto utile per abbreviare i tempi di produzione degli articoli, delle presentazioni e dei poster, soprattutto quando i gruppi di ricerca sono sovranazionali. In questo articolo descriveremo la nostra esperienza nell'uso del software sharelatex basato sul linguaggio di scrittura accademica LaTeX. I prodotti editoriali accademici realizzati con questa piattaforma sono di alta qualità.

Keywords: Scrittura Collaborativa, LaTeX, ShareLaTeX.

1 Introduzione

La collaborazione nella produzione di testi scientifici è una necessità determinata dalla attuale modalità lavorativa dei gruppi di ricerca. Nella maggior parte dei prodotti editoriali accademici, gli autori appartengono a istituzioni scientifiche diverse o a gruppi di ricerca sovranazionali. Appare ovvio che se i ricercatori desiderino collaborare attivamente alla stesura di un prodotto editoriale accademico debbano disporre di strumenti di scrittura collaborativi.

Le tecnologie di internet e della comunicazione hanno ridotto i tempi nella redazione dei testi anche nel caso di gruppi di ricerca sovranazionali. Le nuove tecnologie, note con il termine Web 2.0, consentono ai ricercatori di realizzare prodotti editoriali anche in modalità sincrona, cioè gli autori possono scrivere quasi in contemporanea sullo stesso documento. Attualmente esistono diverse possibilità, sia commerciali che libere, ma tutte si basano sull'uso di un browser di pagine web per la scrittura del documento. Sarà compito di un server remoto coordinare le operazioni di scrittura e di produzione del prodotto editoriale. Una delle scelte da operare è il formato digitale del testo da redigere e conseguentemente anche la tipologia del software di scrittura on line (editor). Queste modalità di scrittura sono indicate con le seguenti sigle:

- **WYSIWYG** (what you see is what you get) ottieni quello che vedi;
- **WYSIWYM** (what you see is what you mean) ottieni quello che intendi;

Nel primo caso l'utente vede immediatamente sullo schermo il documento nel formato con cui sarà stampato; nel secondo l'utente descrive, con un linguaggio di programmazione, come vuole stampare il documento. Nel caso si desideri inviare ai colleghi il documento in formato digitale, il risultato delle due tipologie di editor può essere diverso. Nella prima modalità il documento inviato sarà rielaborato dal software del destinatario e riadattato alle caratteristiche della sua stampante. Si possono così verificare delle modifiche tipografiche del documento non volute dal mittente. Le cause sono la differenza dei formati della carta, dei caratteri tipografici e dei programmi di videoscrittura utilizzati dal mittente e dal destinatario. Con la seconda modalità il documento verrà visualizzato esattamente come ideato dal mittente. Si produce un documento in formato PDF (Portable Document Format) [1] pronto per la stampa su tutte le tipologie di stampanti. Ovviamente i destinatari che desiderino apportare modifiche devono disporre di un programma che interpreti il codice di scrittura del file prodotto dal mittente.

Attualmente esistono programmi per la scrittura collaborativa che si basano su linguaggi HTML (HyperText Markup Language) [2], i quali producono documenti immediatamente visibili online (le pagine web sono le più diffuse) e che si adattano al browser del destinatario (scrittura liquida). In questo caso è praticamente impossibile stabilire a priori la forma finale del documento. Questa modalità di scrittura si limita alla stesura di documenti che devono essere mantenuti sempre aggiornati e sempre disponibili online agli utenti. In questa classe di documenti troviamo guide d'uso, voci enciclopediche, notiziari scientifici.

Nell'articolo proponiamo una piattaforma informatica di scrittura collaborativa online basata sul linguaggio di scrittura LaTeX[3]. Il software è stato valutato sia per le diverse esi-

[a] *Istituto di Struttura della Materia - C.N.R., via Salaria km 29.300, 00015 Monterotondo, Italia*
[b] *Istituto di Cristallografia - C.N.R., via Salaria km 29.300, 00015 Monterotondo, Italia.*
† Rapporto tecnico IC-RM 2016/04 protocollato in data 21/06/2016 n. 0001100

genze di editoria digitale accademica sia per la facilità d'uso nella scrittura collaborativa. Il software, open source, è di produzione della sharelatex[4] ed è disponibile online.[5]

2 LaTeX

Il linguaggio di scrittura LaTeX si basa sul paradigma WYSIWYM. Insieme al contenuto del documento, sono presenti nel testo anche le direttive sulla forma tipografica finale. Con questo linguaggio si possono produrre documenti destinati alla stampa quali: libri, articoli, tesi di laurea, lettere, curriculum, presentazioni e poster.

Vantaggi di questo linguaggio sono:

- L'automazione della composizione tipografica (sommari e indici), inclusa la numerazione capitoli e paragrafi, i riferimenti incrociati, le tabelle e le figure, l'organizzazione delle pagine.
- Ottima resa delle equazioni matematiche e della loro impaginazione.
- Gestione dei riferimenti bibliografici.
- Editor e compilatori sviluppati e distribuiti con licenza open source.

La struttura tipica di un documento LaTeX è la seguente:

```
\documentclass[]{book}
  ...
  preambolo
  ...
\begin{document}
  ...
  testo della pubblicazione
  ...
\end{document}
```

La prima riga identifica la tipologia del documento che deve essere prodotto. Nel nostro esempio un libro. Per questa tipologia l'editor ha un gruppo di specifiche di base su come deve essere impaginato il documento: il formato, i margini di stampa ecc. Segue un preambolo dove sono inserite le richieste di specifiche librerie per il processamento di specifiche parti del testo, quali le figure, le lettere accentate, il correttore ortografico, le equazioni matematiche, simboli speciali. Sempre nel preambolo potranno essere inserite specifiche direttive per realizzare template (modelli) di documenti specifici per i diversi prodotti editoriali. In rete sono disponibili molti modelli per le diverse esigenze editoriali[6]. Molte riviste scientifiche chiedono agli autori di utilizzare i propri template per la scrittura dell'articolo nella forma tipografica finale. Esistono anche dei pacchetti specifici per la rea-lizzazione di diapositive per presentazioni, comunicazioni orali[7] e poster[8].

I comandi $\backslash begin\{document\}$ e $\backslash end\{document\}$ racchiudono il testo del documento. In caso di testi molto lunghi è sempre possibile suddividere il testo in più file e collegarli con il file principale con il comando $\backslash include\{nomefile\}$. Altro vantaggio del programma è la gestione dei riferimenti bibliografici. Attraverso un semplice editor di testo si può creare un file bibtex (ad esempio *bibliografia.bib*), contenente tutte le nostre citazioni bibliografiche, aggiungendo ad esempio il testo esportato dal sito dell'editore della rivista o da google scholar. Qui di seguito un esempio di citazione bibliografica in formato bibtex:

Fig. 1 Area di lavoro per la scrittura del documento con sharelatex.

```
@article{righini2013,
  title={Progetto Calliope: La Piattaforma di
  e-Publishing dell'Area della Ricerca RM 1.},
  author={Righini, Guido and Ianniello, Luca
  and Nantista, Giuseppe and Ricci, Claudio
  and Pifferi, Augusto},
  journal={SMART eLAB},
  volume={1},
  year={2013}
}
```

Con i comandi $\backslash cite\{righini2013\}$, $\backslash bibliography\{bibliografia.bib\}$ e $\backslash bibliographystyle\{unsrt\}$ si creano i riferimenti nel testo e l'elenco delle fonti citate. La numerazione delle citazioni è una delle procedure automatizzate del programma.

3 ShareLaTeX

Di recente sono stati sviluppati editor LaTeX, con interfaccia grafica per gli utenti, che ne facilitasse l'uso. ShareLaTeX è un editor grafico on-line che consente di scrivere in modo collaborativo i documenti in formato LaTeX. Più autori possono operare in contemporanea sullo stesso testo. Un cursore colorato con una etichetta ci mostra dove il nostro collega ha posto il suo cursore. Se il collega inserirà una frase noi la vedremmo comporre in tempo reale sulla nostra pagina web. La pagina web prodotta dal programma è suddivisa in tre sezioni (vedi figura 1). Nella sezione di sinistra sono riportati tutti i file necessari alla creazione del documento: i files con il testo in formato LaTeX, le figure, eventuali fogli di stile (.sty) e il file della bibliografia in formato bibtex. Nel margine in alto del settore di sinistra, le icone per gestire i file (rimuove, aggiungere, rinominare, creare nuove sottocartelle). Nella parte centrale il file da editare. Infine nella sezione di destra l'anteprima del file pdf che si genera compilando il documento. Sul margine in alto a destra della finestra, le icone per compilare, per condividere con i colleghi il progetto editoriale, per vedere la cronologia o i log e per avviare una chat con i colleghi.

4 Discussione

La classe di documenti "article" è quella utilizzata in prevalenza sulle riveste scientifiche, perché consente di realizzare un prodotto editoriale su due colonne con ottima gestione dei riferimenti e delle fonti bibliografiche. Gli articoli presenti in questo numero della rivista Smart eLab, sono stati realizzati in LaTeX con la nostra piattaforma basata sul software sharelatex (http://latex.mlib.cnr.it). La gestione delle figure e delle tabelle all'interno delle colonne è risultato più semplice rispetto all'uso dei word processor tradizionali quali Word (MS Office)

e Write (Libreoffice e Openoffice). La numerazione automatica delle figure, delle tabelle e delle fonti bibliografiche è molto efficiente. Se si inseriscono nel testo, ad esempio una nuova citazione o una nuova figura, tutto sarà di nuovo aggiornato durante la compilazione finale del testo. La resa tipografica del documento finale è molto buona.

Un altro dei prodotti editoriali accademici più realizzati dai ricercatori è la presentazione. La presentazione è una serie di dispositive (slide) che contengono brevi testi, figure, tabelle e equazioni matematiche. Molte di queste presentazioni sono rea-lizzate con PowerPoint o Impress; esse possono contenere anche delle animazioni che accompagnano il discorso dell'oratore durante la presentazione. Entrambi i software precedentemente citati lasciano al ricercatore ampia libertà di realizzazione della presentazione con una interfaccia grafica molto amichevole. Per essere proiettate queste presentazioni, sul pc della conferenza deve essere presente lo stesso software con cui sono state realizzate. La dimensione del file può risultare molto pesante per pc con limitate risorse hardware e in alcuni casi i colori risultano essere diversi da quelli originali. Qualora si desideri una versione cartacea o PDF della presentazione il risultato tipografico può essere deludente, soprattutto in presenza di animazioni. Nel caso di LaTeX esiste un pacchetto di nome **Beamer**[7] che aiuta il ricercatore a creare una presentazione con font chiari e di giusta dimensione direttamente nel formato di stampa PDF. Per proiettare la presentazione basterà attivare la modalità presentazione su lettore di file PDF. Questi lettori sono multi-piattaforma, leggeri, gratis e in alcuni casi opensource. Il pacchetto beamer prevede anche dei comandi per realizzare alcune animazioni, quali apparizioni successive di testo e figure nella stessa diapositiva. Nella figura 2 un esempio di una diapositiva realizzata con LaTeX. Il pacchetto consente di realizzare un modello di diapositiva con tutte le indicazioni sugli autori, la data, le affiliazioni e una filigrana con logo. Inoltre è possibile realizzare dei sommari con link ai paragrafi, citazioni e riferimenti bibliografici.

Utilizzando il pacchetto beamer è anche possibile realizzare dei poster sulla piattaforma on-line basata sul software sharelatex. Attraverso la piattaforma gli autori potranno realizzare, modificare e produrre in modalità collaborativa il file PDF pronto per la stampa in formato A0. In figura 3 un esempio di un poster realizzato in modo collaborativo.

Fig. 2 Esempio di Presentazione realizzata con LaTeX e Beamer.

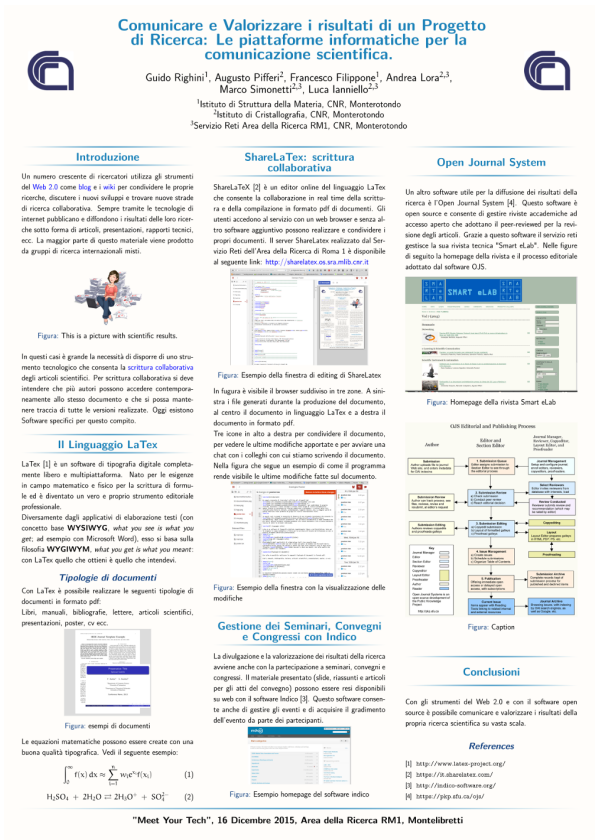


Fig. 3 Esempio di poster realizzato con la piattaforma latex.mlib.cnr.it in formato A0.

5 Conclusioni

La piattaforma informatica basata sul software sharelatex si è dimostrata adatta alla creazione di prodotti editoriali accademici in modalità collaborativa. Tramite la piattaforma gli autori potranno confrontare le diverse versioni del documento creato e discuterne tramite il canale chat. Altri vantaggi sono la semplicità d'uso della piattaforma, la possibilità di vedere e correggere i propri prodotti da pc, connessi alla rete, anche se privi dei software LaTex.

A partire da questo numero della rivista *Smart eLab* tutti gli articoli saranno realizzati con questa piattaforma.

Riferimenti

1 https://it.wikipedia.org/wiki/Portable_Document_Format.
2 https://www.mediawiki.org/wiki/MediaWiki.
3 http://www.latex-project.org/.

4 H. Oswald, J. Allen, Sharelatex sito web: http://sharelatex.com.
5 Sharelatex sito github: https://github.com/sharelatex/sharelatex.
6 http://www.latextemplates.com/.
7 https://bitbucket.org/rivanvx/beamer/wiki/Home.
8 https://www.sharelatex.com/learn/Posters.