Cloud-based computer file transfer service for remote stations connected to narrowband Internet

Alessandro Mancini

IIT Institute, National Research Council of Italy
Via G. Moruzzi, 1 – 56124 Pisa (Italy)
email: { firstname.lastname }@cnr.it

temporary assignment to
Italian National Antarctic Research Program (PNRA)
IT specialist at Concordia Station, Dome C, Antarctica
XV Winter Campaign

Technical Report

Prot. n. DC 2018-2019/161

Abstract. Data transfer is an important task in any scientific activity. Concordia Station in Antarctica does not have a reliable Internet connection and the IT facilities are not easy to use for some of the staff.
The implementation using consolidated technologies, based on web interface, has allowed to make a simple file transfer system much more user-friendly and easy to use.
With this system Concordia staff were able to transmit, easily, digital content to their collaborators in the rest of the world in full autonomy.
Since the system is based on a technological infrastructure managed by staff belonging to National Research facilities, storage and confidentiality of data is also kept without resorting to over-the-top Silicon Valley tech companies.

Keywords: *Efficient file transfer*, *user-friendly UI*, *narrow-band Internet*, *cloud*

Table of contents

1. Introduction

In this document we describe a cloud-based file transfer designed for low speed and unreliable Internet connections.

Concordia Research Station, which opened in 2005, is a French-Italian research facility that was built 3233 m above sea level at a location called Dome C on the Antarctic Plateau, Antarctica. In this particular geographic position (75°06'00.6"S, 123°19'57.9"E), Internet service is only available via satellite connection.

Link is established via the IS-19 geostationary satellite, which is located at 166°E longitude. Today, it is the only one that provides coverage to Concordia Station for bidirectional data transmission. The geographical position imposes, to the antenna used for this type of data connection, a satellite tracking with an elevation angle of about 1.2°. This angle introduces a significant amount of noise, therefore, the connection can not guarantee high transfer speeds and outages can happen at any time.

Fig. 1 Location of Concordia Station in Antarctica (75°06'00.6"S, 123°19'57.9"E)

Concordia staff often needs to send digital contents to their collaborators, unfortunately the IT services do not allow to make easy this operation.

Today there are already many Internet sharing and file transfer services, mostly free, as Dropbox[1], WeTransfer[2] or Google Drive[3] to name the most popular, but the unreliable Internet connection at Concordia makes it difficult to upload large files, waits can be long hours before knowing if files has been successfully uploaded.

The e-mail system does not allow attaching files larger than 1 Mb and the number of messages is limited on a daily basis, so for a researcher who has to send documents or publications it can become complicated to transmit even a few megabytes of data. In fact, the savviest staff during the last summer campaign used the WhatsApp[4] mobile app to send their attachments. Being WhatsApp optimized for working over mobile networks (which may suffer from temporary coverage gaps) it is capable of transmitting images and files in a very efficient way. But it is unacceptable that the Concordia staff should resort to these tricks, not to mention the complication of having to transfer documents from the computer to the mobile phone in order to be able to pass them on to their collaborators.

The file transfer system made for Concordia allows sending files of any size in a simple way. The operation is based on a web interface and recalls the most common services on the Internet.

Uploading process is fast because the files are first transferred to a local server, and then data is transferred to a cloud infrastructure where they are made available via a web link, which can be shared. Finally, an email notification system alerts both the sender and the recipient that the files are available for download. Shared files are kept for a limited period of time and then erased.

To best use the limited bandwidth available, files are compressed before sending and the system can resume a transfer in the event of an unexpected connection loss.

---

[1] **Dropbox** is a file hosting service operated by the American company Dropbox, Inc., headquartered in San Francisco, California, that offers cloud storage, file synchronization, personal cloud, and client software. Website: www.dropbox.com

[2] **WeTransfer** is a file transfer service. WeTransfer is based on Amazons infrastructure and technology; it uses Amazons s3 for storage and for sending files. The core service is free, with more features available for premium accounts. Free users can send files up to 2 GB. Website: www.wetransfer.com

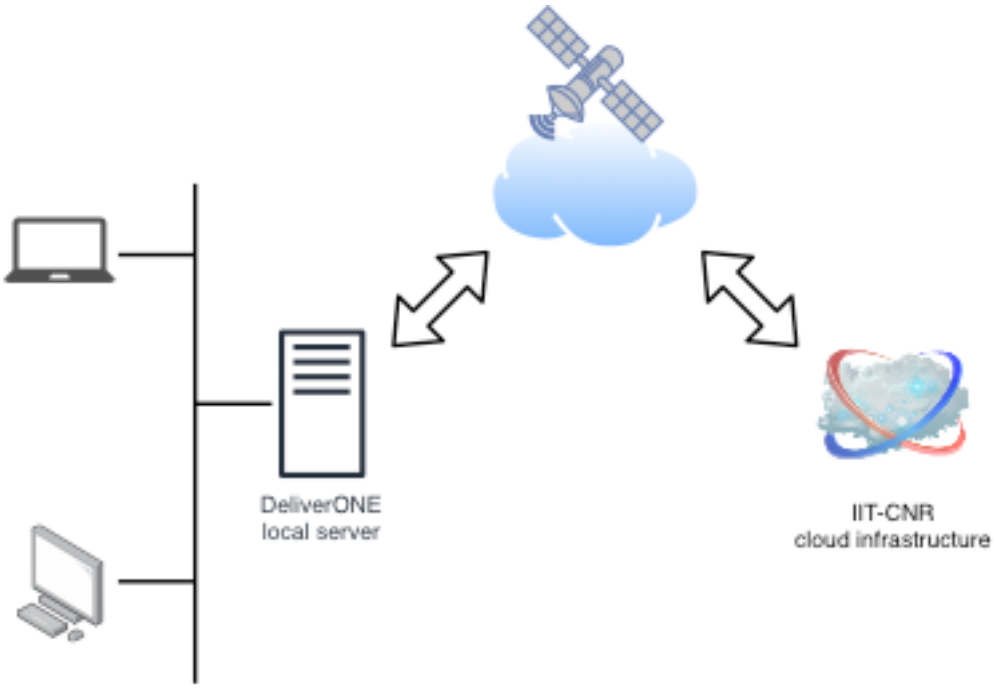
[3] **Google Drive** is a file storage and synchronization service developed by Google, allows users to store files on their servers, synchronize files across devices, and share files. Website: google.com/drive

[4] **WhatsApp Messenger** is a freeware, cross-platform messaging and Voice over IP (VoIP) service owned by Facebook. It allows users to send text messages and voice messages, make voice and video calls, and share images, documents, user locations, and other media. Website: www.whatsapp.com

2. System architecture

System is engineered in two tiers. The main service is provided by the cloud infrastructure at my home institution (IIT CNR[5]) located in Pisa, Italy; while the local server allows users to upload files to be transmitted.

Operating in the Information and Communication Technologies and Computational Sciences field, the IIT Institute have computing and storage resources as well as a strong expertise to manage and maintain IT services internally, without resorting to third-party suppliers, such as Amazon, IBM, Google (to name the most famous). This allows us to store data on our own facility, while keeping privacy and intellectual property under control.
The IIT data center is connected to the Internet via the GARR[6] network and has redundant connections with 20 Gbit/s bandwidth.



The architecture is scalable and several servers, which could be located in different Antarctic stations, can convey data to the cloud infrastructure at the same time.

The local server is NAT aware, it works behind NAT and firewalls because all connections are initiated by itself. Using firewalls it is sufficient to enable only a few communication protocols, as described in the following sections.

[5] Institute of Informatics and Telematics, Nation Research Council of Italy. Website: www.iit.cnr.it

[6] The GARR network interconnects at ultra-high capacity universities, research centres, libraries, museums, schools and other education, science, culture and innovation facilities. Website: www.garr.it

2.1. Network protocols

The operation of the local server is coordinated by the main service, for this reason Internet connectivity is necessary to carry out certain operations.

Messages exchange relies on proprietary APIs[7] that provide cryptographic mechanisms. Each message is encrypted and combined with the server's internal clock: authentication fails if the local server's clock differs for more than 30 seconds from the main service. For this reason it is important to guarantee the reachability of an NTP service to keep the local server clock updated.

The API uses the HTTPS protocol, to guarantee the integrity of the exchanged data while in transit. It is not the purpose of this document to describe the operating mechanisms of the API.

If the local server is protected by a firewall, the following communication protocols must be allowed:

- **dns** – domain name system protocol for address resolution;
- **ntp** – network time protocol, is a networking protocol for clock synchronization between computer;
- **ssh** – secure shell is a cryptographic network protocol used during file transfer
- **ssl** – used by API for secure communication between client and server

If a dns and/or ntp server is already present in the internal network it is possible to avoid this.

3. Service operation

Service is based on web interface. The user does not need to install any applications or key-in complicated commands. The local server can be reached from any computer connected to the local network of Concordia station (both wired and wireless).

The record *deliverone.concordiastation.aq* has been added to the Concordia internal DNS server; it is enough to type the address in the web browser to access the service.

3.1. Service sign-up

Each Concordia guest has an email address *@concordiastation.aq*. The first time the service is used, the user must register, this allows the administrator to keep track of the volume of data transferred and block any misuse.

This operation is very simple and requires the user to enter his name and a recovery email address (different from @concordiastation.aq, Fig. 2). The system immediately sends by

[7] Application Programming Interface (API) is an interface or communication protocol between a client and a server intended to simplify the building of client-side software.

email the password to log-in, without having to ask an administrator (which may not be available).

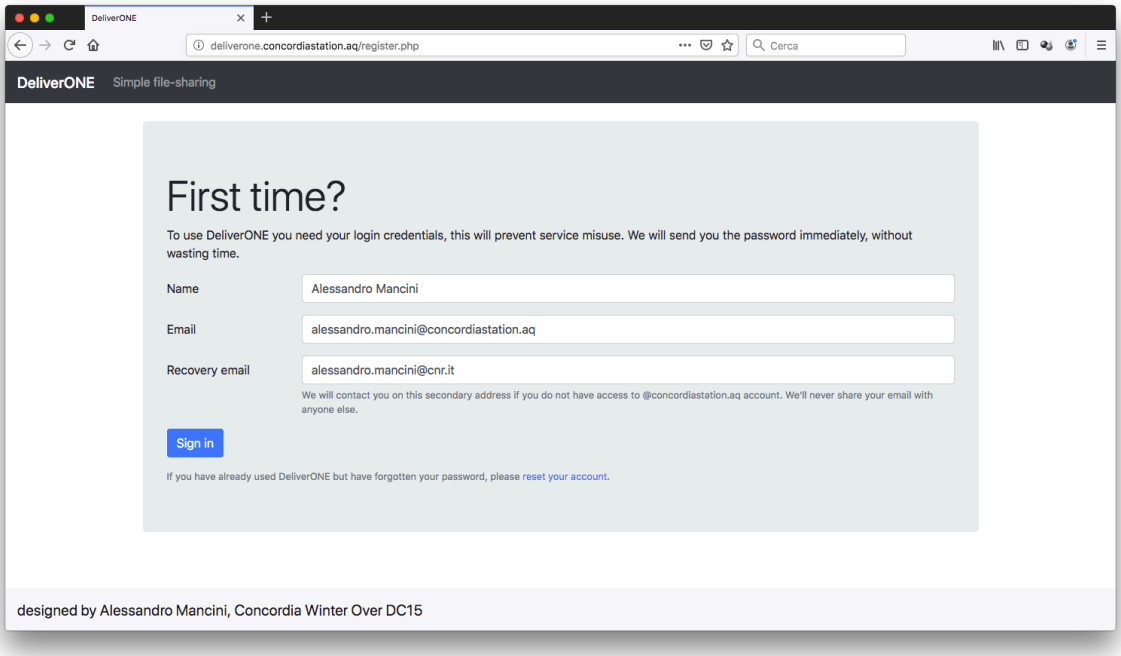


Fig. 2 Service sign-up

3.2. File transfer

To send files, the user only needs to enter the email address of the recipient. In fact, the system transfers automatically data on a webserver and notifies the recipient via email when files are ready for download.

It is possible to add a message to be displayed in the notification e-mail and multiple addresses can be specified for a single transfer (in this way, bandwidth utilization is managed more efficiently).

One or more files can be uploaded via the web interface. If a file is loaded by mistake it can be removed (by clicking on the delete button, see Fig. 3) before confirming the transfer.

At the moment only some types are allowed, allowed extensions are:

- archive (.dmg, .gz, .rar, .tar, .tgz, .zip)
- documents and text (.doc, .docx, .pdf, .txt)
- multimedia (.m4a, .mp4, .mov)
- pictures and photos (.jpeg, .jpg, .png, .tif)
- presentation (.ppt, .pptx)
- spreadsheets (.xls, .xlsx)

If files to transmit are not of the type indicated above, it is possible to create a .tar or .zip archive and transmit that.

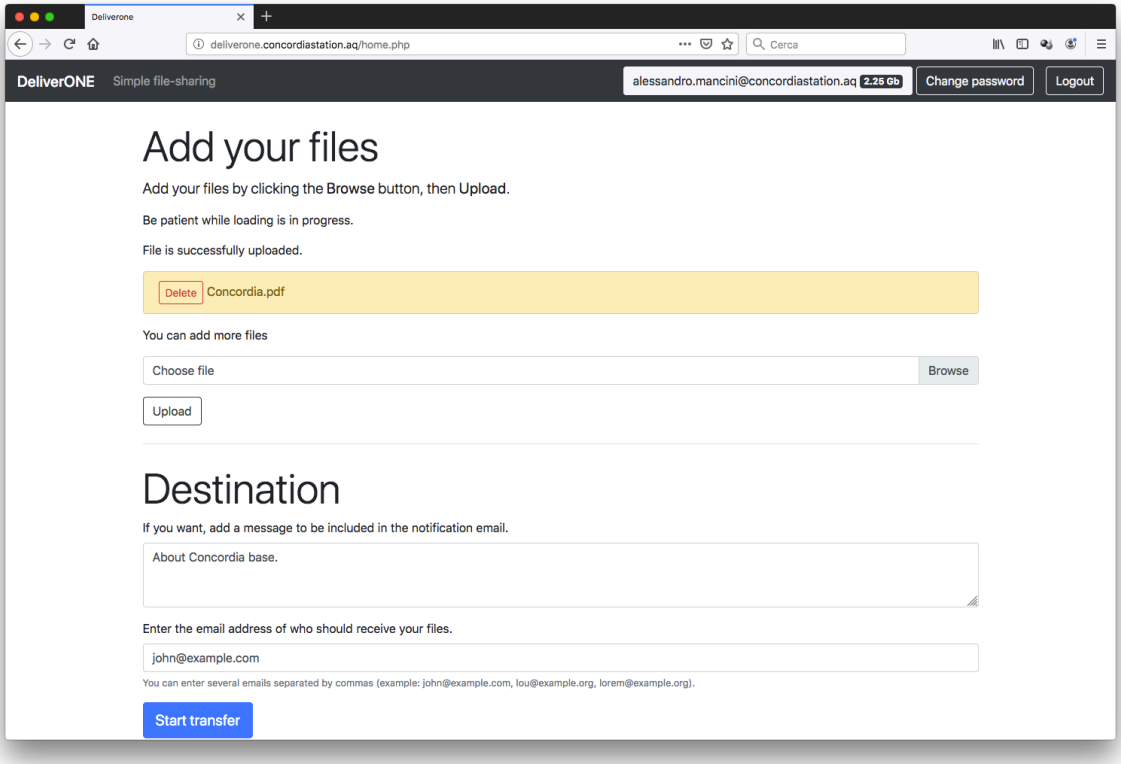


Fig. 3 File transfer preparation

File upload works from a mobile device – smartphone or tablet – as well. This makes easy to send digital content without having to transfer it to a computer. Pictures or documents can be uploaded straight from the smartphone (Fig. 4).

3.3. Transfer notification and download

Once files to be sent are confirmed, the system will start to transfer data to the main cloud infrastructure. This operation could take a long time. As soon as files are available for download, the recipients will be notified by email (Fig. 5).

The e-mail message contains a web link that points to a directory. By clicking on the link, the recipient can easily start download. The web link is generated with a number of random characters, difficult to guess, therefore only those they know the link have access to the files.

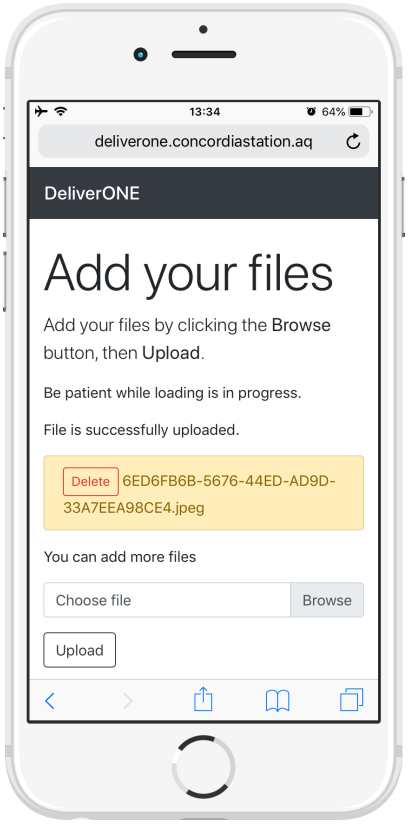


Fig. 4 Picture upload from smartphone

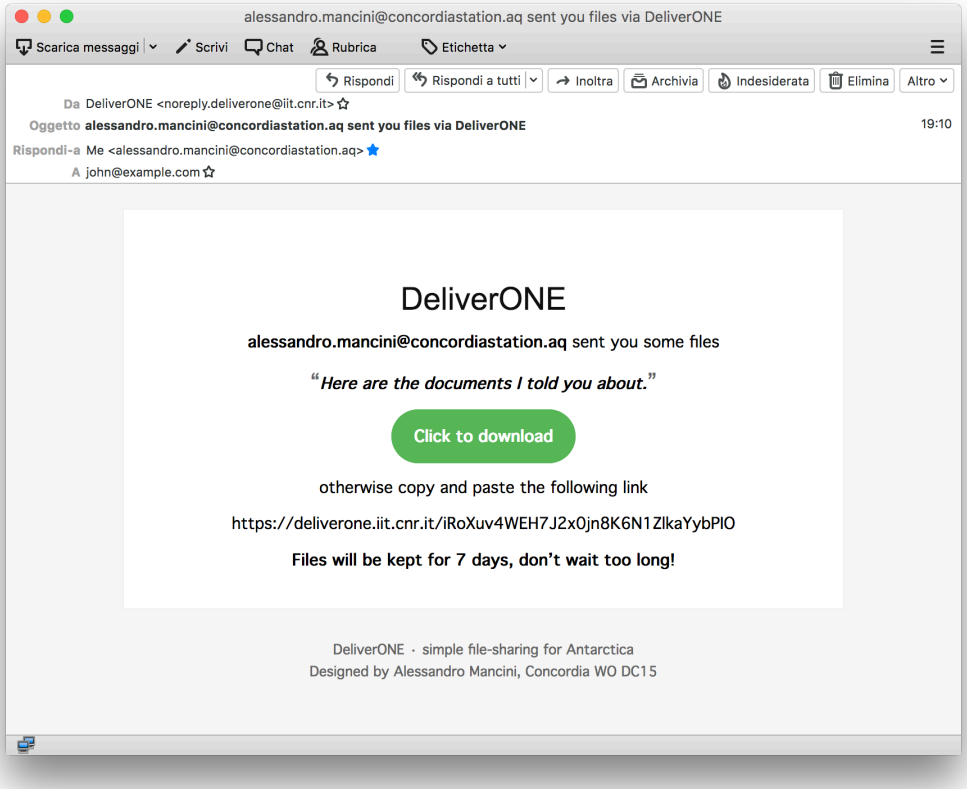


Fig. 5 Notification e-mail sent to the recipient

A confirmation message is also sent to the sender when data have been completely transferred (Fig. 6).

To view and download the files, the recipient simply has to click on the link shown in the message, or copy and paste it in the browser address bar (Fig. 7).

Files are kept for a limited period of time. After 7 days the directory and its contents will be automatically erased from the cloud and will no longer available.

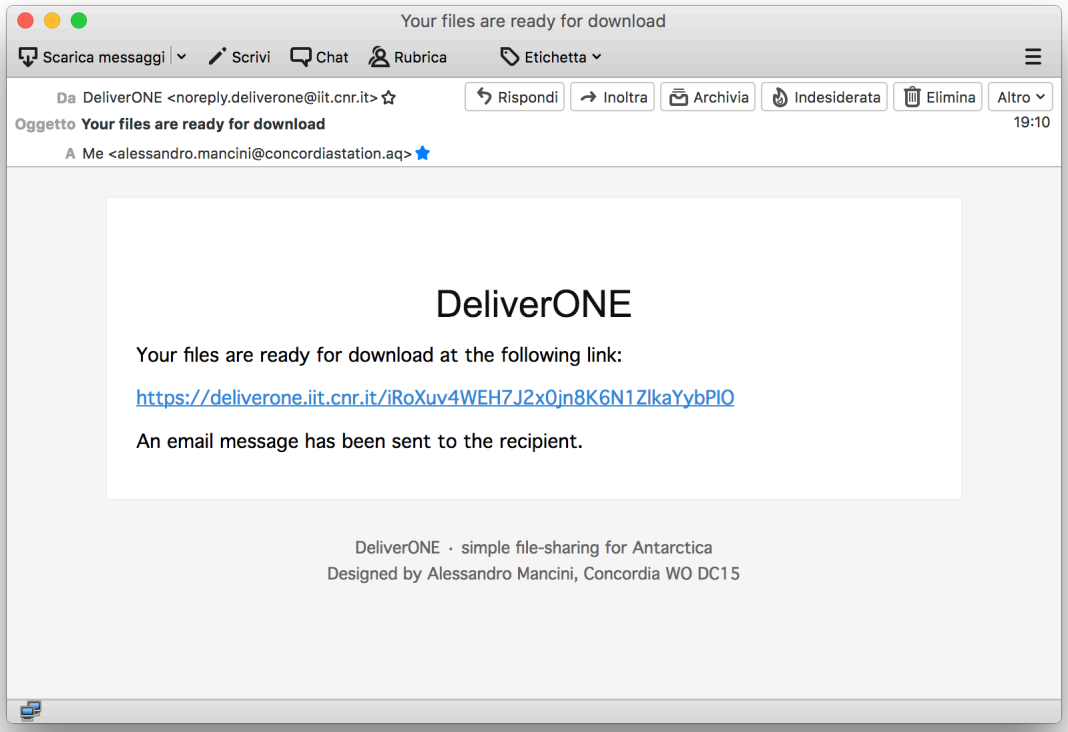


Fig. 6 Confirmation e-mail sent to the sender

4. Administration tools

Service is designed to make users independent and minimize administrator intrusion. Simple tools are available to check system functionality; these can be developed in a more complete way in the future.

These functions are available from the shell.

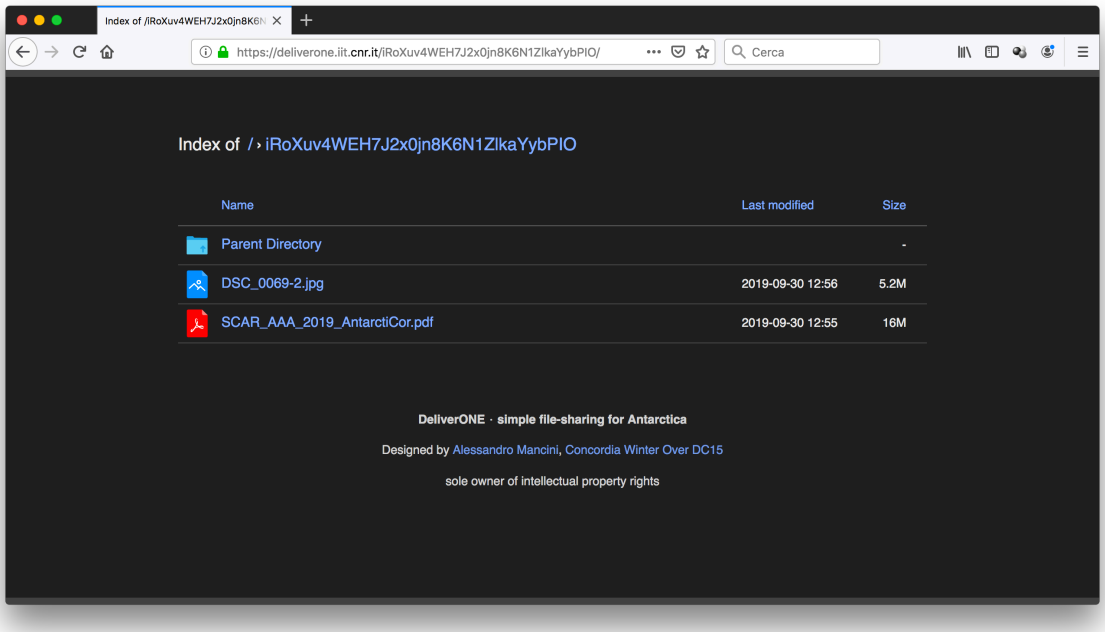


Fig. 7 Download link

4.1. Transmitted data volume counter

This counter shows volume of data transmitted for each account. This function is useful to check which users make use of the service and if the data volume can exceed limits imposed by a code of conduct.

Exmple:

```
root@deliverone:/etc/deliverone/admin# php data-volume.php
alessandro.mancini@concordiastation.aq          2.27 Gb
giuditta.celli@concordiastation.aq              270.81 Mb
julien.moye@concordiastation.aq                 28.52 Mb
massimiliano.catricala@concordiastation.aq      688.65 Mb
nadja.albertsen@concordiastation.aq             1.53 Gb
```

4.2. User password update

The user password is stored in the database using a hash method, so it is not possible to trace or change the old password. It is only possible to overwrite with a new one.

Even without the help of the administrator, the user is able to change or get a new password if login credentials are lost.

Example:

```
root@deliverone:/etc/deliverone/admin# php update-password.php username
```

5. Transfer techniques

To push the narrowband medium more efficiently, files are compressed with bzip2[8] algorithm. This also reduces transmission times; files are then unpacked at destination.

If transmission is interrupted (for example a temporary disruption of the satellite service), the system attempts repeatedly to restore transfer at intervals of time. The transmission will resume without having to retransmit data already sent.

I do not go deep describing the techniques used to prevent being copied without the recognition of intellectual work.

The software architecture allows to easily changing transfer protocols with more efficient techniques, which can be developed later.

6. Future developments

This tool has been developed to meet the needs of winter-over staff during the XV winter campaign at Concordia Station. In fact, the IT infrastructure of Concordia does not allow the use of simple systems for sending files.
The e-mail service does not allow attachments larger than a few hundred kilobytes and the most common file sharing services available on the Internet are unusable. This caused frustration to the staff that needed to transmit a few pages of a publication containing graphics or images.

There is a service called Hermes[9], but it is not for everyone. The system was considered complicated and the user's computer must be equipped with an *sftp* client in order to transfer files to the server. Furthermore, having access credentials is complicated by the ICT staff, which requires forms to be filled out and issues credentials after a long time. Furthermore, the Hermes service during the XV winter campaign was not always available.

All the staff for its ease of use appreciated the new file transfer service. All the users were able to initialize their account and immediately start transmitting files without the help of the administrator. Even the recipients less accustomed to computer technology have never had difficulty downloading data sent to them. This has been possible thanks to the use of consolidated and well-known technologies, such as the web browser, present on all computers without having to install additional software.

If this work can be considered useful for Concordia or other Antarctic stations, the author is open to develop new features to improve the service. The system already allows multiple

[8] **bzip2** is a free and open-source file compression program that uses the Burrows–Wheeler algorithm.

[9] **Hermes** is a file transfer service first implemented by Antonio Litterio (winter-over DC9) and re-engineered by Paride Legovini (winter-over DC10).

installations (which may be present in different stations) to be used simultaneously for sending data on the main cloud infrastructure.
This work can be used freely by anyone; the source code and intellectual work is exclusive property of the author. Software updates can be available in a precompiled code or Virtual Machine ready for installation.

New features, ready to be implemented are:

- **Web interface for administrators**. Now the functions described in section 4 are available only from the shell;
- **File reception**. Users will be able to receive files in the same way described in section 3.3;
- **Quota management**. Allow users to only transmit a certain volume of data per week or per month;
- **End-to-end encryption**. An additional level of confidentiality by encrypting data on server. Only users with the encryption key will be able to decode the contents of a download.

7. Contacts

For more information, follow the status of the project or contact the author, visit **www.deliverone.it**.