Fast: Flexible Automated Streaming transfer tool

Index

Fast: Flexible Automated Streaming transfer tool 1

Index 2

1. Introduction 3
2. Start to transfer an experiment 4
3. Edit an experiment transfer 6
4. Monitor the Rock Samples and Start the Creep phase. 8
5. Cancel and Restart Rock Samples transfers. 8
6. Behind the Fast tool. 11

Attachment (I) 13

Attachment (II) 14

Attachment (III) 15

Attachment (IV) 16

1. Introduction

EFFORT is a multi-disciplinary collaboration between Geoscientists, Rock Physicists and Computer Scientist. Brittle failure of the crust is likely to play a key role in controlling the timing of a range of geophysical hazards, such as volcanic eruptions, yet the predictability of brittle failure is unknown. Our aim is to provide a facility for developing and testing models to forecast brittle failure in experimental and natural data. Model testing is performed in real-time, verifiably prospective mode, in order to avoid selection biases that are possible in retrospective analyses.

The project will ultimately quantify the predictability of brittle failure, and how this predictability scales from simple, controlled laboratory conditions to the complex, uncontrolled real world. Experimental data are collected from controlled laboratory experiments which includes data from the UCL Laboratory and from Creep2 project which will undertake experiments in a deep-sea laboratory.

Depending on how the datasets belonging to an experiment are written in local files, data security policies, and other features, different transfer protocols should be used to upload data, e.g. FTP, SFTP, GRIDFTP, Rsync, Globus Online, etc. In EFFORT project, we have designed a new automated data transfer tool called FAST (Flexible Automated Streaming Transfer) to upload experimental data periodically from UCL or deep-sea laboratories to a RP repository located in a Edinburgh machine.

FAST is compatible with all O.S and it will select the most suitable transfer strategy in each case for allowing to stream data directly to the Edinburgh RP repository. Before starting to upload experimental data, the RP researcher who wants to initiates the data transfer of an experiment need to provide the next information:

- Which channel are involved in the experiment
- Rock Samples Numbers involved in the experiment
- Parent Path
- Type of Experiment
- Type of data transfer
- Location
- Interval of time
- Time out
- Comments
- Type of data
- Contact details

With all these information, FAST will be able to configure the best transfer strategy, and it will upload data automatically to the Edinburgh RP repository.

Besides, metadata about the experiment and the Rock Samples that belong to the same experiment are automatically created and stored in a MYSQL database located in the Edinburgh machine.

If any problem occurs during the data transfer associated with one (or more) Rock Sample(s), FAST will stop the data transfer of this (or these) Rock Sample(s). Furthermore, FAST allows to edit some features of an experiment that is being transferred.

FAST tool has four different tabs that allows to:

- Start to transfer an experiment
- Edit the metadata of an experiment
- Monitor and start creep phase of each rock sample involved in the current data transfer
- Stop/Restart a rock sample transfer.

As follows the main features of FAST will be explained:

2. Start to transfer an experiment

The first tab of FAST is called "New Experiment Transfer". This tab allows to the user to indicate some features of the experiment that wants to transfer to the Edinburgh RP repository.

First of all, the user (RP researcher) has to specify which channels are involved in the experiment, checking the corresponding check boxes. One experiment can involve more than one channel.

After checking the desired channels, the user has to indicated the "Sample Numbers". After writing each Sample Number, press the <enter> key is required. The user can navigate through the "Path information" tree, situated to the right of the tab, to check if the data that have been selected are correct.

Once that the user has selected the channels and indicated the Rock Sample Numbers, the type of experiment can be choose. There are two types: "Real" for transferring a real experiment, or "Trial" for transferring a non-real experiment.

The next feature allows to FAST to ultimate the transfer strategy:

- Type of data transfer: There are three types of data transfer that can be chosen:
    - Transfer an incremental file/minute. New file /day (default): This option allows to synchronize periodically one file per day and per Rock Sample involved in the experiment. For each interval of time, FAST only sends to the RP repository the differences between the previous transfer and the current file. At the end of the day, FAST moves the local file that has been synchronized to another local directory.
    - Transfer new file /minute: This option allows to transfer periodically a new file per Rock Sample and per interval of time. In this option, for each interval of time, FAST sends a new file to the RP repository, omitting the files that have been sent previously.
    - Backup: This option allows to make a remote backup of an experiment. It is a single transfer operation (not periodically), that

copies the Local Rock Samples (selected) folders and their contents to the RP repository.

- Location: Could be London (default) or Catania.
- Interval of time: This is the period of time between new data is stored in the local machine. 1 minute by default.
- Time out: This is the maximum time that Edinburgh server can wait until give up because any new data have arrived in this period of time. 8 minutes by default
- Parent Directory: This is the path of the parent directory which contains CH1, CH2, CH3, CH4, CH5 and CH6 directories. Inside of each CH are located the subfolders Sample-NUMBER. Inside one of those subfolder should be the files which contains the data to transfer: "strain-timestamp-.txt"
    - For example, if the user wants to transfer a new experiment to the RP repository with CH1 selected and the Number of Rock Sample 0, FAST will search for data in the following path:
        - PARENT-DIRECTORY/CH1/Sample0/strain-timestamp-txt
    - The Parent directory can be configured in the profile file (transfer.properties) or the user can select a new one pressing "Select Path".

The user can also write comments about the experiment that is going to be transferred. By default, the type of data that are going to be transferred are Stress, Strain and AE. However, the user can deselect any of these types. Finally, some contact details about the user who wants to start to transfer an experiment are required. In the following tabs, all the information is filtered by the user name specified here.

Once all these data are filled, the user can press the "Start transfer Data" button. In that moment the metadata about the experiment will be sent to database located in Edinburgh machine, and the data will periodically [1]upload to RP repository located as well in Edinburgh.

If the user wants to reset values of this tab, only has to press "Clean Fields" button. Currently we are working on storing and selecting different experiment configurations. But in this moment this option is not ready yet. So, if the user press the "Transfer Configuration" button, nothing is going to happen

The Figure 1 shows the first tab of FAST that we have already explained.

---

[1] Transfer an incremental file, and transfer new file options, are periodically data transfer operations. However, Backup is a single data transfer operation.

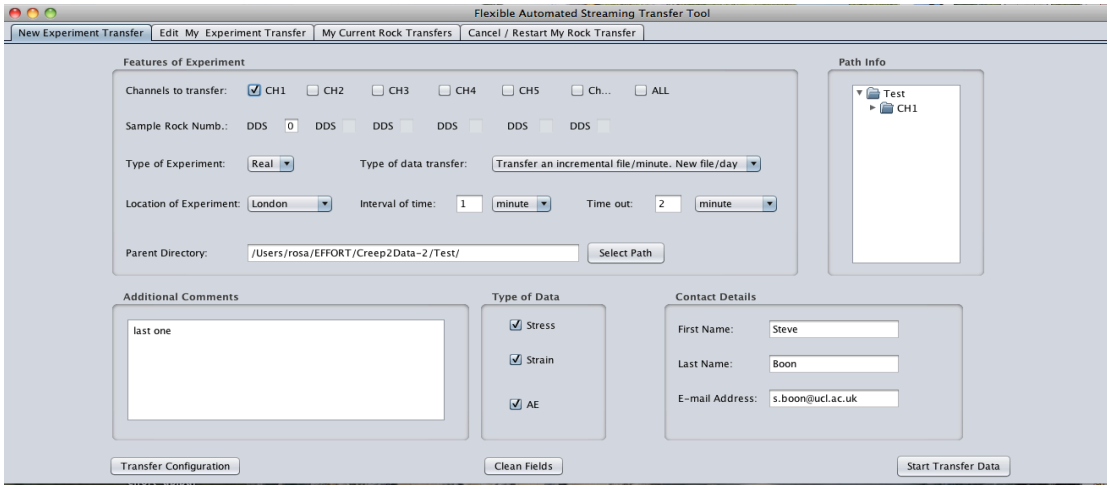


Figure 1. Start to transfer data

3. Edit an experiment transfer

The second tab of FAST is called "Edit My Experiment Transfer". This tab allows to the user to see and edit some metadata related with his experiments.

As we introduce before, every time that the user starts to transfer data from one experiment, metadata is automatically sent to a Mysql database located in a Edinburgh machine. This database is composed by two tables called Experiment and Rock:

- Experiment . In this table, the metadata of each experiment is stored. This table contains the following fields:
  - ID_Experiment: Unique for each Experiment.
  - Timestamp_Start: Time stamp when the experiment has started
  - Time_Interval: Interval of time before arriving each new data
  - Time_Experiment: Duration of the experiment
  - Type_Data: Experimental, Sinthetic.
  - Notes: Comments
  - Owner_Data: Contact details
  - Location: London or Catania
  - Metadata_Path: Path of the Metadata file in the remote machine.
  - TimeOut: Period of time that the server can wait without any data.
  - Type_transfer: Incremental file, new file, backup.
  - Stress: True or False
  - Strain: True or False
  - Ae: True or False

- Rock. In this table, the metadata of each Rock Sample is stored. This table is linked with Experiment table, because one Experiment can contains one or more Rocks. The Rock table contains the following fields:
  - ID_Rock: Unique for each Rock.
  - Timestamp_Finish: Time stamp when the rock has failed or its data transfer has been stopped.
  - Timestamp_Creep: Time stamp when the creep phase has started

- Complete:
  - 0: If data associated with this rock have been transferred
  - 1: If the rock has failed
  - 2: If the user has stopped the data transfer by pressing STOP button
  - 3: If the server has wait the "Time Out" and any new data has arrived .
  - 4: If suddenly the Local Rock Sample directory is empty
- Path: Remote path where the data has been stored
- Channel: Channel of this Rock.
- Append_Path: Remote append path
- ID_Experiment: Id_Experiment associated with this Rock.

Notice that ID_Experiment and ID_Rock could be different values. For example, imagine that the last ID_Experiment was 27, and the last ID_Rock was 28. If now a user wants to transfer a new experiment which contains 3 Rock Samples, the ID_experiment will be 28, and each rock will have ID_Rock 28, 29 and 30.

After giving this short description about the remote database and its tables, Figure 2 can be better understood.

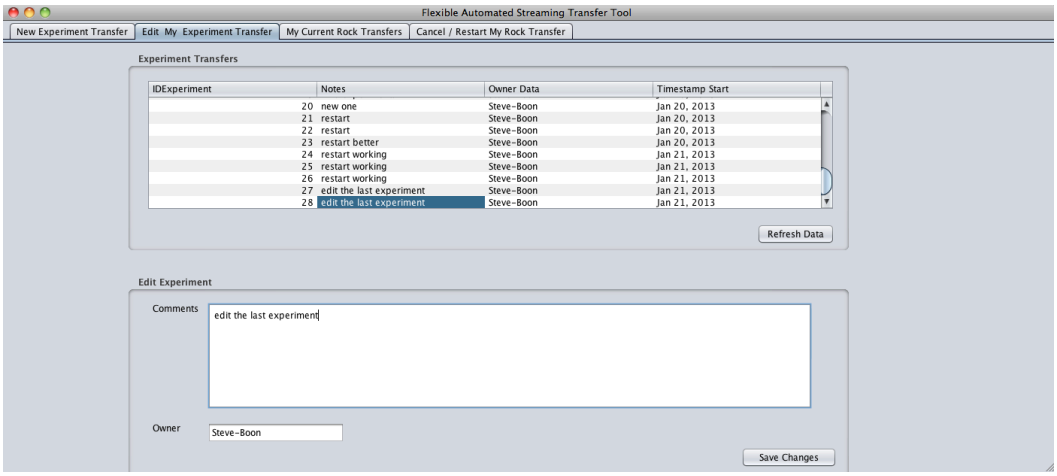


Figure 2. Edit an experiment with ID_Rock 28.

In the top of this tab, we can see a table called "Experiment Transfers". This table shows some fields of the Experiment table (remember: the Experiment table is located in a Edinburgh machine, no in the users' computer) that we have explained before, like: ID_Experiment, Notes, Owner Data, Timestamp Start. Note that the user only can see its experiments in this table.

If the user selects one of the rows of this table, he can edit the comments and owner name by using the text fields that are in the bottom of the tab. As soon as the user presses the "Save changes" button, the changed will be automatically saved in the remote database, specifically in the Experiment table.

4. Monitor the Rock Samples and Start the Creep phase.

The third tab of FAST is called “My Current Rock Transfers”. This tab allows to the user to check the data transfers that currently are being transferred. Once a Rock sample has failed or stopped, its associated information will be disappear from this tab.

At the top of this , some features of the Rock Samples that are being transferred are showed in a table called “Current Rocks Transfers”. If the user selects one rock sample(clicking with the mouse), he can check if the data transfer is working well by reading the LogFile associated with that Rock Sample. Also, the user can check the size and timestamp of the files of that Rock Sample in the remote machine.

Another thing that this tab allows is to initiate the Creep Phase of a Rock Sample. The user only has to select one Rock Sample from the table and press the ”Start Creep” button.

In Figure 3 are showed the option that we have already explained.

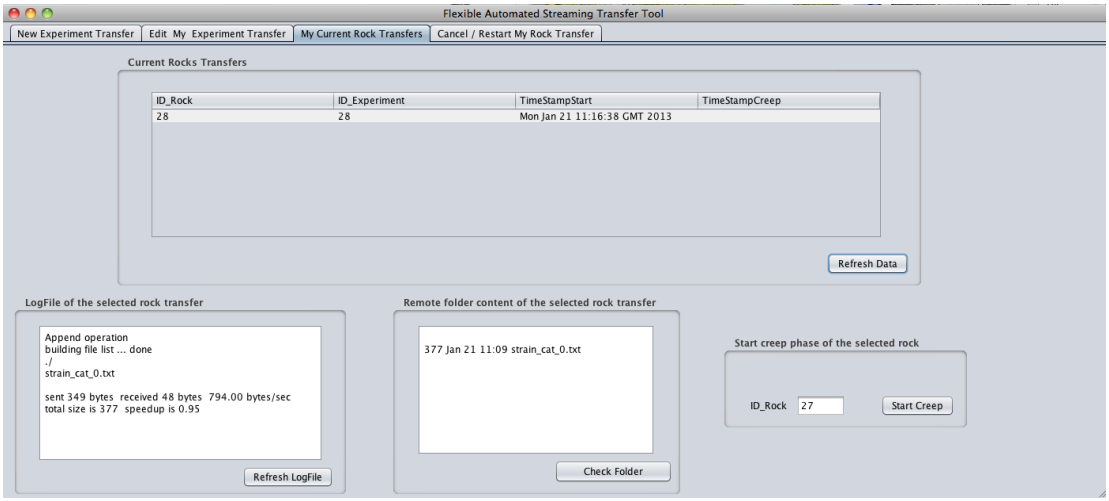


Figure 3. Check the current Rock Transfers

5. Cancel and Restart Rock Samples transfers.

The last tab of FAST tool is called “Cancel/Restart My Rock Transfers”. At the tab name indicates, this tab allows to select one of the data transfer that are being transferred to stop it. Or restart it in case that it has being stopped by one of this three causes:

- The user has stopped the data transfer by pressing the stop button

- The server doesn't receive new data. Time out has expired.
- Suddenly, the Local Rock Sample folder is empty.

At the top of this tab a table called "Rock Data" is showed. This table is a join between the Experiment and Rock tables (located in the Edinburgh database).

Is quite important to remark the possible values of the field Complete:

- 0: If data associated with this Rock Sample has been transferred
- 1: If the Rock Sample has failed
- 2: If the user has stopped the data transfer by pressing STOP button
- 3: If the server has wait the "Time Out" and any new data has arrived .
- 4: If suddenly the Local Rock Sample directory is empty

So, a user can only to stop the data transfer associated with a Rock Sample, if its completed field value is "0".

In case that the user wants to restart a data transfer, he can only do it if the selected Rock Sample complete field value is "2", or "3", or "4".

In the Figures 4, 5, 6 and 7 show features described before

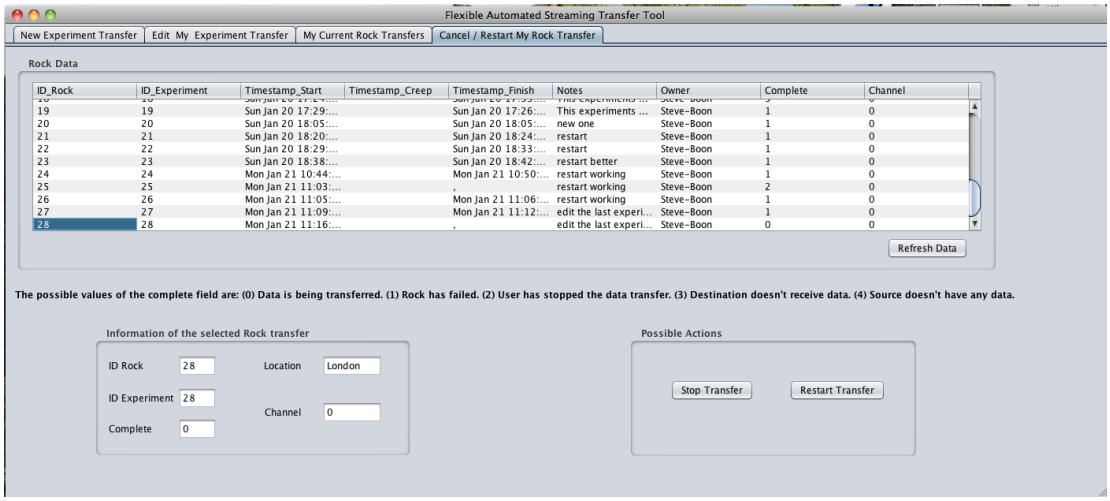


Figure 4. Selecting Rock 28 to stop the data transfer

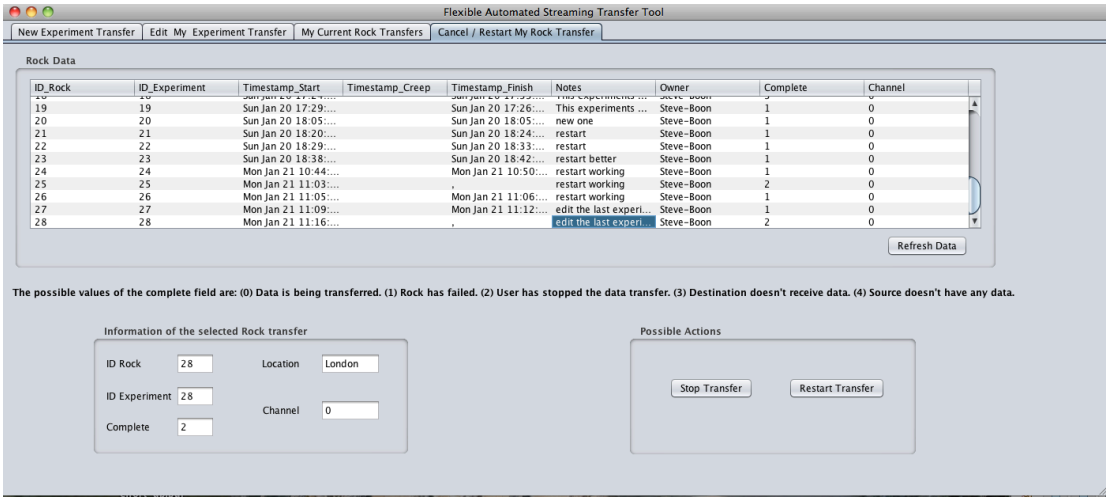


Figure 5. After stopping Rock 28

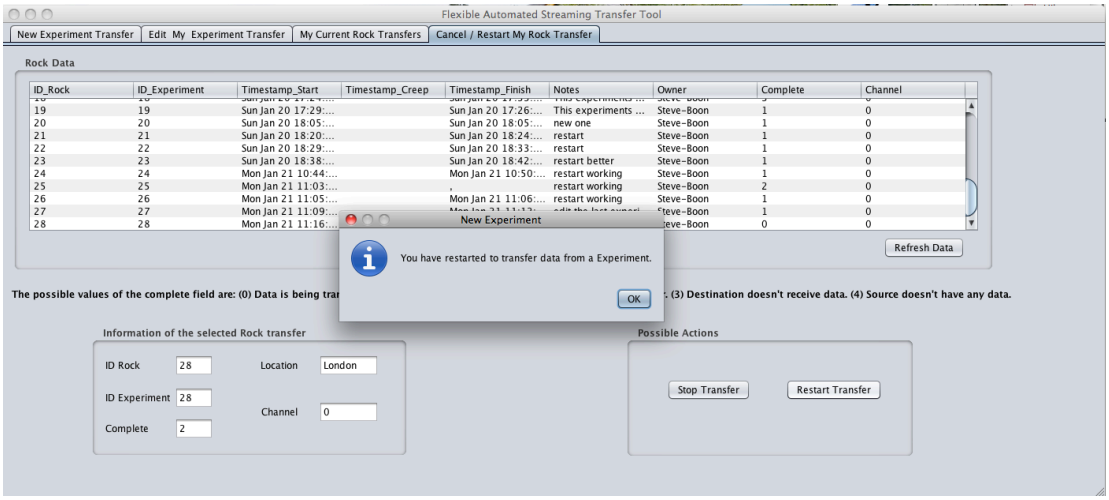


Figure 6. Selecting Rock 28 to restart data transfer

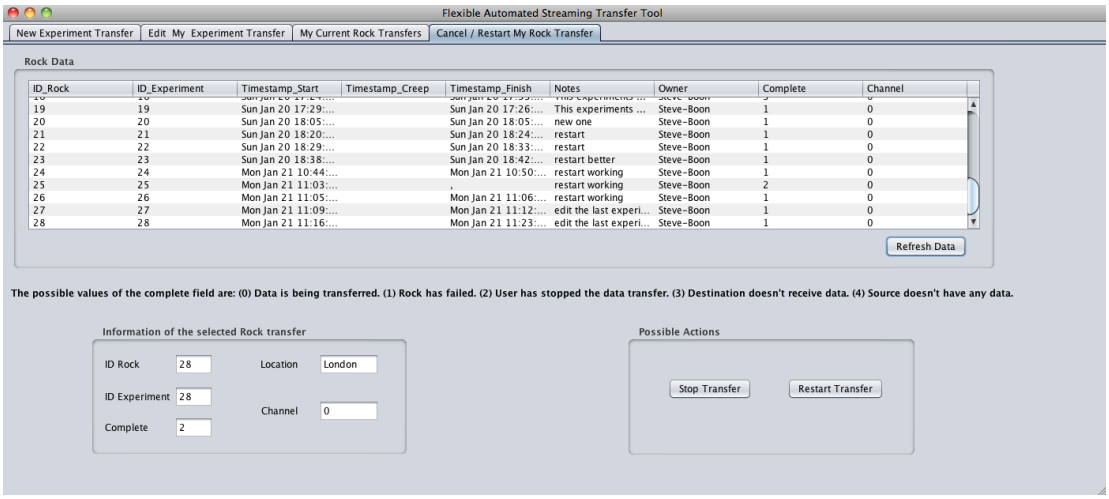


Figure 7. After restarting the data transfer. Rock 28.

Notice that if suddenly the local computer lose the network connection, or the local computer has log off, after the time out, the server will change the status of complete to "3" to all the data transfer operation that has being performed.

So, if the user wants to restart all the data transfers of all the Rock Samples involved in the same experiment, the user has to restart them one by one. In other words, FAST allows to restart a Rock Sample and not an experiment.

Finally, FAST will notice that a Rock Sample has failed if in its Local Rock Sample folder is stored a file called "end.txt".

6. Behind the Fast tool.

Behind the Fast tool there are several programs that runs automatically. Some of them are in the local machine [2](transfer-channels.sh, stop-channels.sh, restart-channels.sh) and others in the remote machine (receive.py, append.py, finish.py, re-receive.py).

When FAST tool is installed in the user machine, a copy of transfer-channels.sh, stop-channels.sh, restart-channels.sh are stored in the local machine.

Also three directories are created in the same directory where FAST has been installed:

- -LOGS: In this directory FAST stores some local files like: Information-IdRock.txt, SentFiles-IdRock.txt, LogFile-IdRock.txt, Error.txt, output-timestamp.txt, stop-IdRock.txt.
- -METADATA: metadata-timestamp.txt
- -etc: transfer.properties

The only file that needs to be explained carefully is the transfer.properties. The other files are created and deleted automatically by transfer-channels.sh, stop-channels.sh and restart-channels.sh programs.

Once that user has installed the FAST tool in a directory inside his user's computer, he has to enter into this installation directory, and open the "FAST-INSTALLATION/etc/transfer.properties" file to edit it with next information.

---

[2] Local machine is the user's machine or computer. Remote machine is the Edinburgh machine or server.

User has to specify the following path:

```
bash_path = Is the path of the “bash” in your computer.

script_path= Is the path where the transfer-channels.sh, stop-
channels.sh, restart-channels.sh are stored. Normally, in the
same directory where FAST has been installed.

data_path = Is the parent directory. Inside this directory
will be CH1, CH2, etc channels. Inside each channel directory
will be stored SampleNumber subdirectories.

parent_path = Should be the same that the previous one.
```

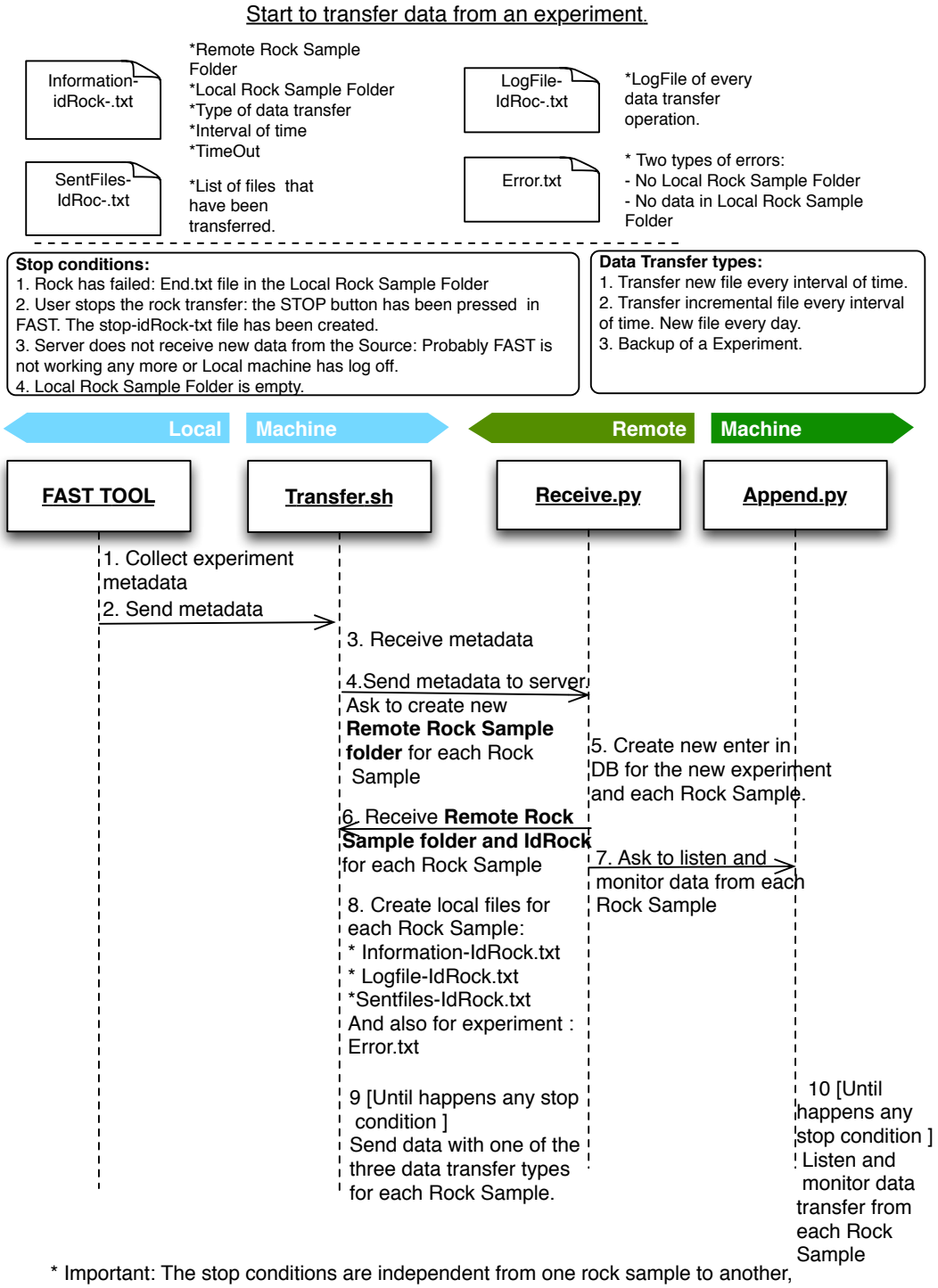
As follows there is an example of the previous values:

```
bash_path = /bin/bash

script_path= /Users/rosa/Desktop/FAST/

data_path = /Users/rosa/EFFORT/Creep2Data-2/Test/

parent_path = /Users/rosa/EFFORT/ Creep2Data-2/Test/
```

In order to clarify how the FAST tool works behind Guided User Interface, some graphics are attached to this manual. In those graphics, we can see how the different local and remote programs interact among themselves.

As well, a graph of the Experiment and Rock tables of the database is attached to the end of this manual. In this graph we can see the relation between both tables. Creep_Data is only a view of those two tables. In other words, the view is a selection of some fields of both tables in a unique table.

Attachment (I)

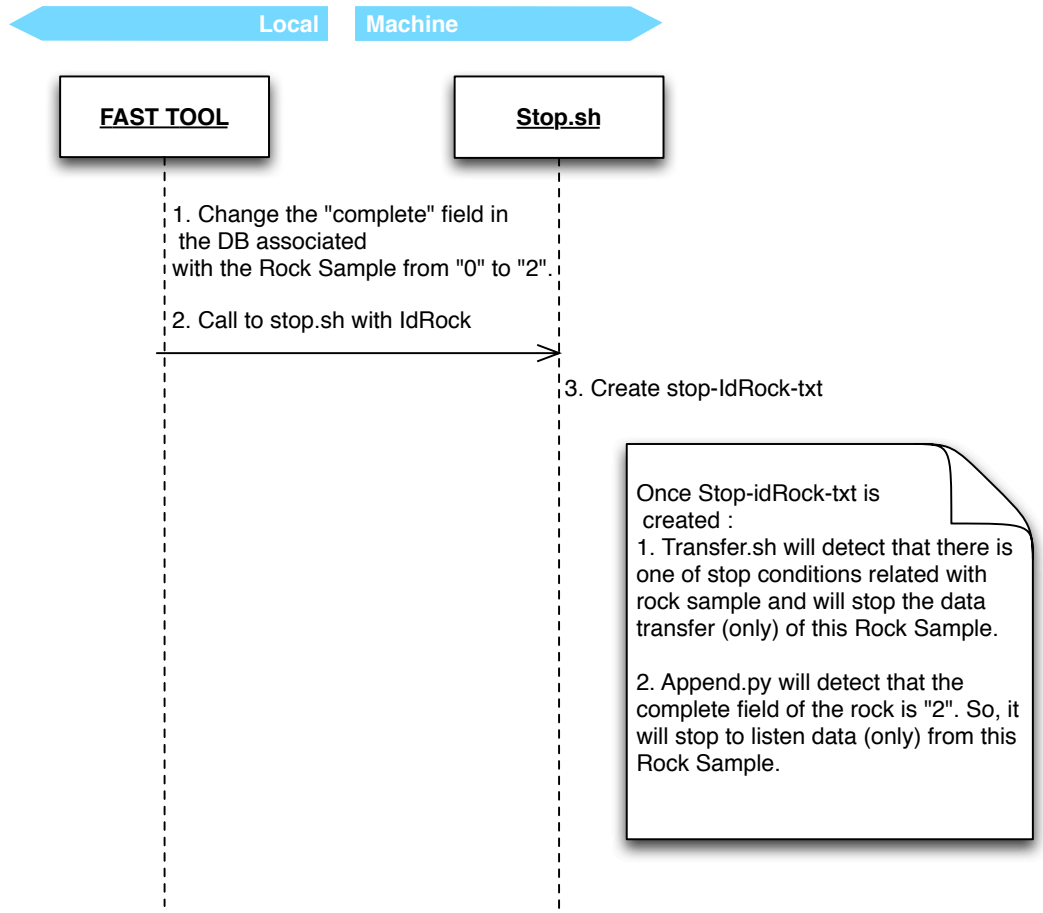


* Important: The stop conditions are independent from one rock sample to another, although they belong to the same experiment.

Attachment (II)

Stop a data transfer associate with a Rock Sample

The user has selected through FAST tool a Rock Sample that want to stop to transfer. The rest of Rock Samples associated with the same experiment will be transferring normaly.



Attachment (III)

Restart to transfer a Sample Rock

Information-idRock-.txt

*Remote Rock Sample Folder
*Local Rock Sample Folder
*Type of data transfer
*Interval of time
*TimeOut

LogFile-IdRoc-.txt

*LogFile of every data transfer operation.

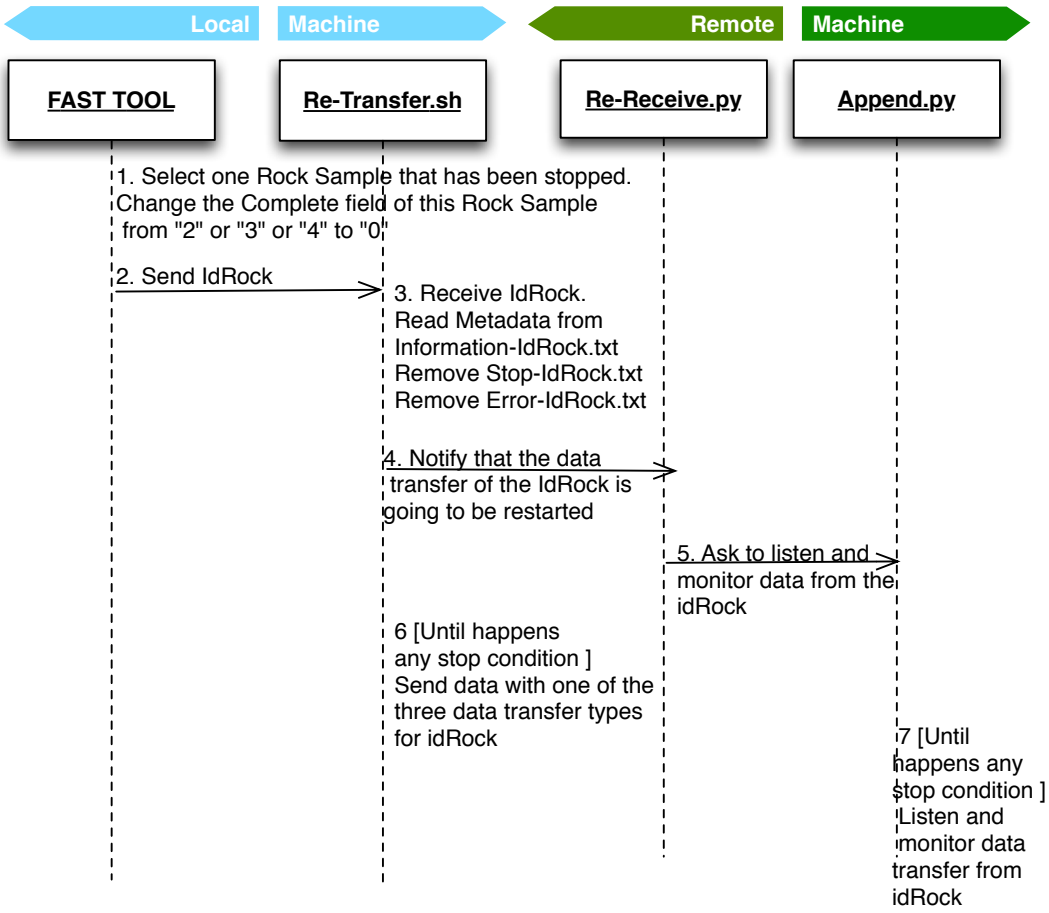
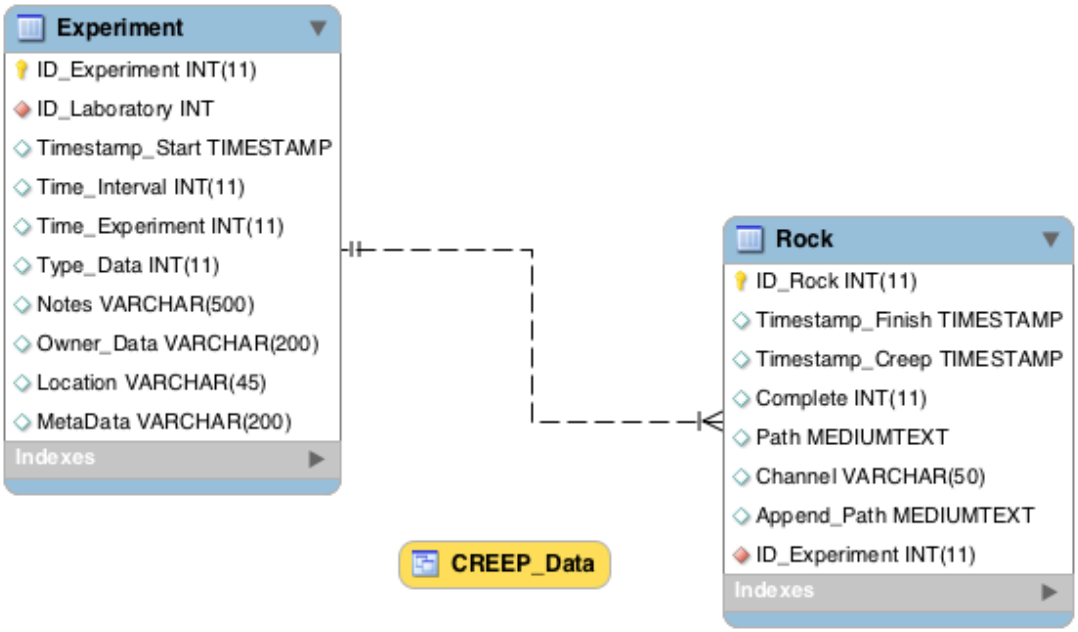
SentFiles-IdRoc-.txt

*List of files that have been transferred.

Error.txt

* Two types of errors:
- No Local Rock Sample Folder
- No data in Local Rock Sample Folder

The user selects through FAST tool one of the Rock Samples that has been stopped.
Restart to transfer only affects to the Rock Sample that the user has selected.
If more than one Rock Samples that belong to the same experiment have been stopped, the user has to restart them one by one.



Attachment (IV)