Plan Overview

A Data Management Plan created using DMPonline

Title: Exploding Wire Open Data Plan

Creator: Gonzalo Rodríguez Prieto

Principal Investigator: Gonzalo Rodríguez Prieto

Data Manager: Gonzalo Rodríguez Prieto

Affiliation: Other

Template: DCC Template

ORCID iD: 0000-0001-6302-0146

Project abstract:

Exploding wire research following these main topics:
- Exploration of metals in gas state and its characteristics.
- Electrical discharge in atmosphere.
- Long time plasma research.

ID: 80702

Start date: 06-09-2021

End date: 01-07-2099

Last modified: 27-06-2022

Exploding Wire Open Data Plan

Data Collection

What data will you collect or create?

Data created and collected from this project are:

- Signals relative to electrical and other kind of probes whose raw exit is a voltage collected with an oscilloscope.
- Transformation into the adequate units of the previous signals.
- Photos of the oscilloscope screen images
- Photos and espectra obtained from multiframe, streak, and spectrometer devices available in the laboratory.

All of them are expected to be in digital formats, listed below. In case that some experimental data will be a physical object, a digital counterpart will be stored, when possible.

Digital formats used are:

- CSV, TXT file format for the oscilloscope data, both the raw and transformed values.
- JPG for the photos of the scope channels.
- Proprietary formats, TIFF and DAT text images for the streak camera images.
- Multi frame camera produces raw photos in standard RAW format, and also saves the same pictures transformed to more common formats. Currently chosen are 16 bit TIF and JPG, with the addition of a XML file with the shot setup information and a JPG file with a thumbnail association of all the photos.
- Spectrometer produces three files, one in a proprietary format, and by later manual processing a CSV file and a TIF image of the spectrum.

Expected volume of raw data per shot is 0.28 GB. An usual experimental campaign produces ~100 experiments, so expected volume of data per campaign is 28 GB.
In order to share the data, due to high volume, only the fraction of them experimentally significant will be publicly shared on the Zenovo repository.
Nevertheless, all the data are available upon request by e-mail. Transfer and share of the data will be deal with as soon as the need arise.
To store permanently this data, USB hard disks with proper labels are used.

How will the data be collected or created?

Procedure of data collection:
Data collection is performed in these separate steps:

1. First, experiments are running and the necessary devices are expecting the experimental trigger, upon which their control computers or units (for oscilloscopes, mainly) present the data in screens.
2. Data from the devices are stored into their hard disks computers, and oscilloscopes data and screenshots are stored in a dedicated computer upon calling a custom created software manually by the researcher in charge of the experiments.
3. Hand recollection of all the data from the used experimental devices in their computers is made, at least, once per year, to transfer the data to the dedicated hard disk for permanent storage.

Naming and folders convention:
The following table is actualized as of 2022-March. On it, "LLXXX" identifies the shot following one of two conventions: For the first 1000 experiments, three numbers, eg. "123". For the next experiments, two letters followed by three numbers: "AA123".
From 2022-March on, any changes on the naming convention will be here written to, so actual used convention and interpretation is here available.

Device	Naming and folders convention
Oscilloscopes	Folder called "ALEX-Elec" has all the shots data. Inside, each shot has a dedicated folder called "ALEX-LLXXX-elec". Inside this folder, some files with basic information over the shot is stored, later explained, and at least another directory called "RAW". Inside this RAW folder channels voltage raw trace without any modification are stored in CSV format files. Convention names for these raw channels is that first appears the scope followed by the channel name. For example: "ALL-CH1.CSV". Frequently it appears other directory called "WORKED" that contains CSV files with the scaled and transformed voltage traces into measured quantities. The group of files called "XXXX-picture.bmp" have screenshots of the scopes, with "XXXX" the name of the oscilloscope, and are stored in the main folder for each shot, together with the file named "ALEX-LLNNN.html" with basic information over the experiment.
Hamamatsu streak camera	Folder called "ALEX-Streak", were all the experimnt files are stored together. File name convention for each shot or experiment: *ALEX-LLXXX-LLNN.ITEX* Proprietary version of the file data. "-LLNN" makes reference to the camera sweep time. It contains a lot of metadata about the Streak and camera setup, timing, etc. *ALEX-LLXX-LLNN.DAT* DAT format, consisting on a matrix with the intensity values of the image. *ALEX-LLXX-LLNN.TIF* A TIF image with the same information that the previous *.DAT file.
Cordin, an image converter device.	Folder called "ALEX-Cordin". Inside, each shot has a dedicated folder called "ALEX-LLXXX-Cordin-YYYY-MM-DD", with the structure "YYYY-MM-DD" the date of the experiment in the format year, month, day. Each shot folder contains the following files and folders: 1) "Capture.XML", an automatic recollection from the software controlling the camera of the device settings. 2) "thumbnail.jpg", a fast vision of all the cameras information. 3) folder "16bppTIF": A folder with a 16 bites per point recollection of the converter images in TIF format 4) folder "AverageJPG": A folder with a JPG average images of the converter images. 5) folder "RAW": A folder with the images in raw format. All the internal folders follow the naming convention: "FrameXX_FORMAT.FFF", with FORMAT.FFF" dependent on the type of image: "16bpp.TIF", "low.JPG" and "2048x2048.raw" for the three mentioned folders, respectively.
Fanta, a spectrometer	Folder called "ALEX-Fanta". Different format files following the convention "ALEX-LLXXX-YYYY-MM-DD.FFF" with "FFF" changing following the format of the data and "YYYY-MM-DD" the year, month and date of the experiment, are all together into the main data folder for this device. Data format are the propietary "spe" form the spectrometer and "CSV" open comma separated values for wavelength and intensity.

Documentation and Metadata

What documentation and metadata will accompany the data?

There is a register of every shot made, in a ODS libreoffice datasheet format file, with the most important information ordered by column. This information includes: date, wire material, length and diameter, voltage in the capacitors, which devices are on, and the reason for the shot. Its name is "ALEX-AA000--Shot-List.ods" for the last shots recorded.

Currently, there is a clear division on stored data between the "electrical" data, which are from different probes attached to scopes and any other device data.

For the electrical data, channel identification, order and reason for the shot with some minor comments about it are stored into the HTLM file within the folder with the raw scopes data, always stored, and other folder with different data treatments.

Concerning other kind of experimental devices, currently there are three:

1. Hamamatsu streak camera
2. Cordin multiframe system
3. Fanta: Spectrometer with an iCCD attached.

All of them have dedicated computers, so the data are stored in these dedicated computers in specific folders. Metadata with information about experimental conditions and setup of these devices is generated and stored automatically by their own control software when the format is the propietary of the devices. In order to access to this information, control programs of the devices are currently necessary.

First storage of the experimental data is done in folders dedicated for each device within the pertinent computer of control, as previously explained. For the Hamamatsu, folder is called "Streak data\ALEX streak", for the Cordin "ALEX-Cordin", for the spectrometer, "ALEX-Fanta", and for the oscilloscopes, "ALEX-elec".

Hamamatsu streak camera.

As explained before, these data are stored into three formats: first, the proprietary from the company, then a TIFF file and third a text file with (*.DAT) that contains the pixels intensity values in a matrix. Metadata into the proprietary file format includes time of recording and setup characteristics of the experiment.

Cordin multiframe system:
Their files are stored into a folder named with the shot name and data (ALEX-LLNNN-Cordin-YYYY-MM-DD format. "LLNNN" the naming convention for the shots, "YYYY-MM-DD", the date in year, month, day format) with sub-folder for the distinct formats, that can vary from shot to shot. Always the raw photographs are stored and a XML, "Capture.xml", file with timing and CCD gains, exposure times and other setup information is stored.

Fanta spectrometer:
First file stored is with the proprietary format data with the image of the camera, extension "spe", and then different, open formats of these image: usually, the spectrum obtained by summation of all the columns and a graphical image of the camera vision. All share the same name, and only the file termination indicates its type (CSV for the spectrum, JPG or similar for the image and SPE for the proprietary one). As commented, name convention is "ALEX-LLXXX-YYYY-MM-DD.Extension". Proprietary format data includes time and data of the shot, with information over software version, hardware attached to the computer, timing characteristics and other relevant information.

All these data are stored after the shot has finished and the operator manually orders so. The file with the shots information is fill up manually after some group of related experiments shots.

Ethics and Legal Compliance

How will you manage any ethical issues?

There are no ethical issues to consider in these data, as they are just pure physical experiments.

How will you manage copyright and Intellectual Property Rights (IPR) issues?

Copyright and ownership of the created data belong to the University of Castilla-la Mancha, as the work is performed in installations of its property.
Data are made available if used in experiments and scientific works, with some months delay and perhaps an embargo considering scientific article publication, through the used of public data repositories for data. Currently, Zenodo web page.
When data are of tests and other experimentation, not interesting for the scientific community, they are also openly available by request to the data responsible.
Employed software is open and licensed by Creative Commons licenses when possible. This open software is available at the GitHub web under alfeliz user (https://github.com/alfeliz)

Not restrictions on availability or use of data are considered on the long term, but for plagiarism or other misconducts and any ordered embargo or prohibition of working with people or institutions of certain countries by the adequate government bodies.

Storage and Backup

How will the data be stored and backed up during the research?

Data storage is made during experiments on the computers that control the devices directly, on the folders mentioned before, immediately after the experiment.

A manual back up process is made, at least, once per year by the data responsible. In this process, all the folders with the data for each device are copied on a dedicated hard disk. This copy is used as base for another copy on a working computer, to have two separate copies of all the raw data. Up to now the hard disk size increase over time is way beyond the data generation velocity. Such situation will remain in the near and medium term future.

In the case of any incident implying data loss, they will be copied back again to the device control computers.

How will you manage access and security?

Data stored in these experiments are not sensitive in any way, as no traded secrets or personal data are necessary on this research. Main risk to data security is the simultaneous failure of the backup computers and hard disk, but such possibility is remote enough as not to worry about.

Therefore, to maintain secure the data, back up systems are stored in adequate locations inside the laboratory, to which only the main researcher and closed collaborators have access.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

Data to be retained are essentially all the experiments made with the ALEX system. There is no plan to destroy or delete any data obtained from the experiments, including the data that are of non experimental use. Due to the low storage needs for all this data, in the foreseeable future the needs of storage will be far behind the possibilities offered by the technology. Therefore, preservation of the data is hoped to continue, at least, until retirement of the responsible of the data.
Main research uses of this stored data are now two:

1. Validation of novel research results and
2. The use of electrical signals to better understand metal phase space.

What is the long-term preservation plan for the dataset?

All data are available directly by request to the data manager, and from year 2021, some selection of them used in published research are being made available in the repository Zelondo as a specific collection for each published article.
The data stored in the laboratory does not incur in any storage cost, out of the price of hard disk and associated hardware, readily paid by the host institution. Data formats and folders are selected directly during the experiments, and later maintained unaltered for the storage. Due to the use during this period of multiple format file saving, with at least one format open and free, no problems are thought to happen in the data sharing.

Data Sharing

How will you share the data?

All data obtained from this experiment are available to anybody who request access to them. Only exceptions are embargoes to certain countries or institutions mandated by my own country or higher order institutions, like European Union.
As the total dateset is diverse and very difficult to show by any electronic means to a single person, in order to access them an e-mail request is necessary. Such e-mail must include the reasons for data accessing and which specific data are necessary. If the information provided is not enough to clarify either the needs of the requester or necessary data, further explanation will be provided in the answer to the e-mail, so data to be send, its nature, and approximate number will be clearly known by the data receiver in advance.

The method to share the data will depend on the amount of data. So, if it is small enough as to not overload any e-mail provider, this will be the preferred method. When the amount of data is too large, then the use of the university of Castilla-la Mancha cloud system with an individual access to the request person will be preferred.
All the data will be made available in the shortest possible time, with a compromise of answering the e-mail request within 72 hours. Data availability will follow answer and previous accordance on their nature and quantity, and within two weeks from the agreement.
Although an identifier of the data is recommended in general, when ALEX stored data are requested, it is not mandatory to pursue one. Data employed directly in research is available through the open data repository Zenodo. For example, raw experimental data on resistivity measurements for iron and platinum are at the web: https://zenodo.org/record/3648264#.Yk7FhDyxVrs

Are any restrictions on data sharing required?

No restrictions on data sharing are expected, out of the previously signaled imposed restrictions by the Spanish government or

higher order bodies.

Responsibilities and Resources

Who will be responsible for data management?

There is only one person responsible for all the data management from the ALEX experimental system, including data capture, metadata production, etc. The responsible for this whole plan is Gonzalo Rodríguez Prieto, who also writes it.

What resources will you require to deliver your plan?

Planned resources for this plan were already previously described, and includes hardware in form of hard disks to copy the data form the acquiring computers to their definite location in the USB hard disks.