# Metadata for Charisma

## Introduction

The need of Charisma

## Scope of the project

What are we doing and why

## State f the art on Raman protocols, standards accepted and regulations.

Refer to literature and presentations WIP at WP3

### Source of norms, calibrtions and standards used in Raman

ISO norms related to Raman Spectrosocpy [ref]

ASTM norms involving Raman [ref]

Pharmacopeia regulations for Raman, US, EU, Jp, Ch, others [ref]

Other regulations.

### Source of materials available for Raman

NIST for intensity correction

NIST for spectrometer calibration (emission lines) [ref], spectrometr companies supplying Ne, Ar, He light source for this purposes. White light for intensity calibration, (source)

Jap reference for Si [ref]

Elodiz new created reference samples for Raman [ref]calcite, si and Polystyrene

Others

## Definitions on terms and approach on classification on areas of control

We will follow the advice from the OSTER project (EU grant 760827)

The definition of steps is perfectly applicable to CHARISMA and the harmonisation of Raman data; 4 elements to be separated:

* Instrumentation and its control –method of analysis
* Raw Data
* User case –sampling
* Data processing

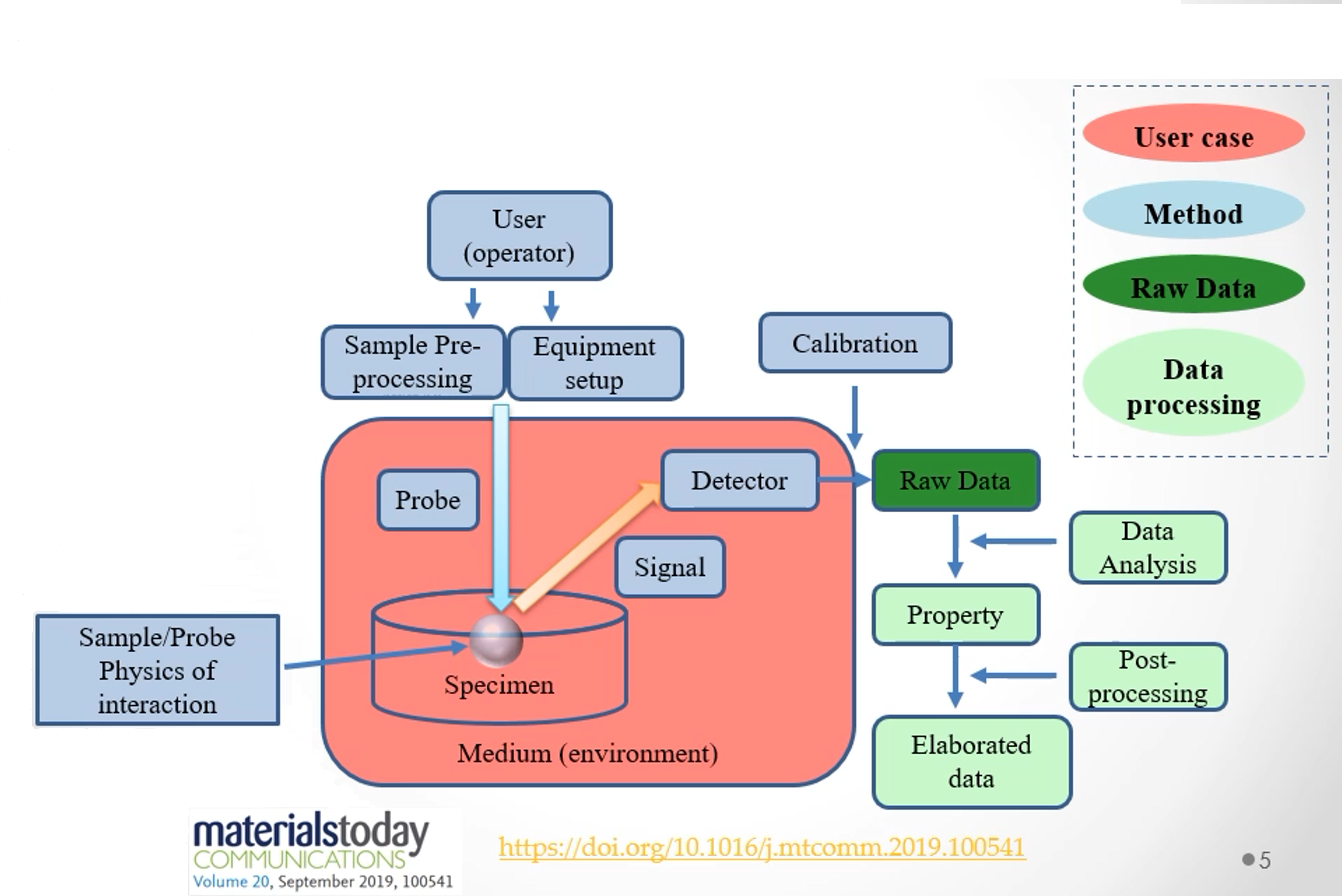


Fig1. OYSTER project, classification areas for data acquisition and analysis. Orange, user case, refers to sampling, how the sample will be analyse and the environment/ecosystem required to do so, blue refers to method, defined as the hardware required to run the experiment. Dark green refes to Raw data, as generated from the sample and produced the analytical tool, prior any processing on the data, other that he necessary to produce meaningful data. AND LIGHT GREEN, refers to data processing, that is the data manipulation of make the data useful and adapted to purposes.

### Instrumentation and control. Method of analysis -method-

Some very well stablish methods already in place to verify these units –reviews prepared in WP3 by elodiz Athens, CSIC, frouwahve- and some areas that CHARISMA has already identified as top priority to develop tools and techniques for harmonisation.

We have separated the hardware requirement from type of instrumentation/performance/flexibility and capability to adjust/modify the data. From usability perspective, we can classify the Raman units in 3 groups:

* + 1. Raman microscopes, high-end systems where user has unlimited possible combination of configurations and optical path/setups. Includes hyphenation and specialise Raman techniques. Data is heavily affected by the hardware setup. Each optical path should have a verification and qualification specific protocol. There are often easily recalibrated and reconfigured. uncorrected Data -raw data- is in general accessible.

These devices have a very large variability of configurations and they are mostly used for research purposes where that flexibility is required. These unts are as well the main cause of discrepancies on the results published, as the reproducibility from unit to unit and between manufacturer interpretations introduce the most discrepancies on data [ref and example]

Example: graphene samples using confocal vs portable data.

* + 1. Medium range of Raman setup, unit is with a limited number of configurations, this includes modular setups. These units are very often without any moving parts and qualifications and calibrations are permanent or highly stable. The number of optical paths and setups is limited, and the qualification of the device is simpler, but often, it is difficult to change protocols for qualification and recalibration by the users. Data generated is often pre-process and corrected by the unit firmware and not easily access to raw data.
    2. Raman analysers, units that use the Raman spectroscopy in a closed environment and users have very limited or no capability to change the system and where data is fully processed. Customers are unlikely to have access to raw data.

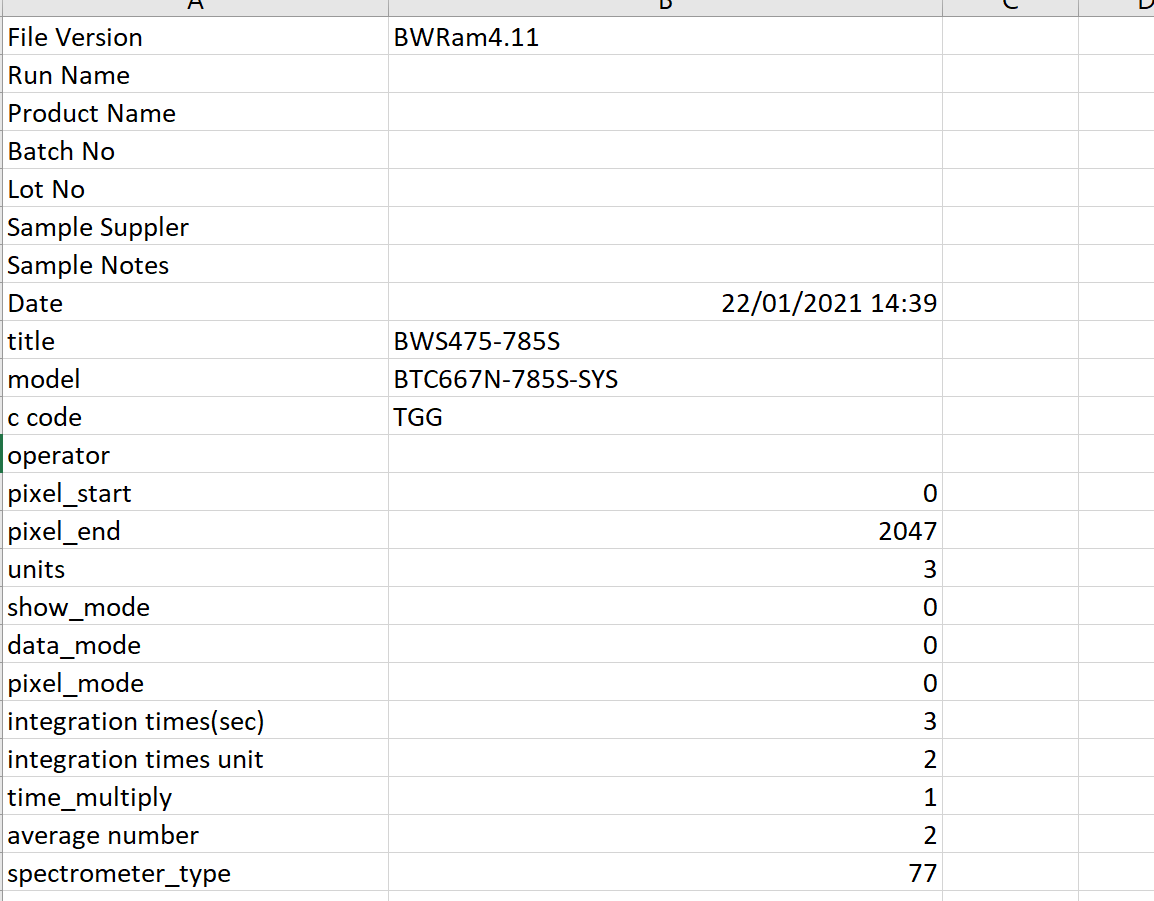
Raman instruments and

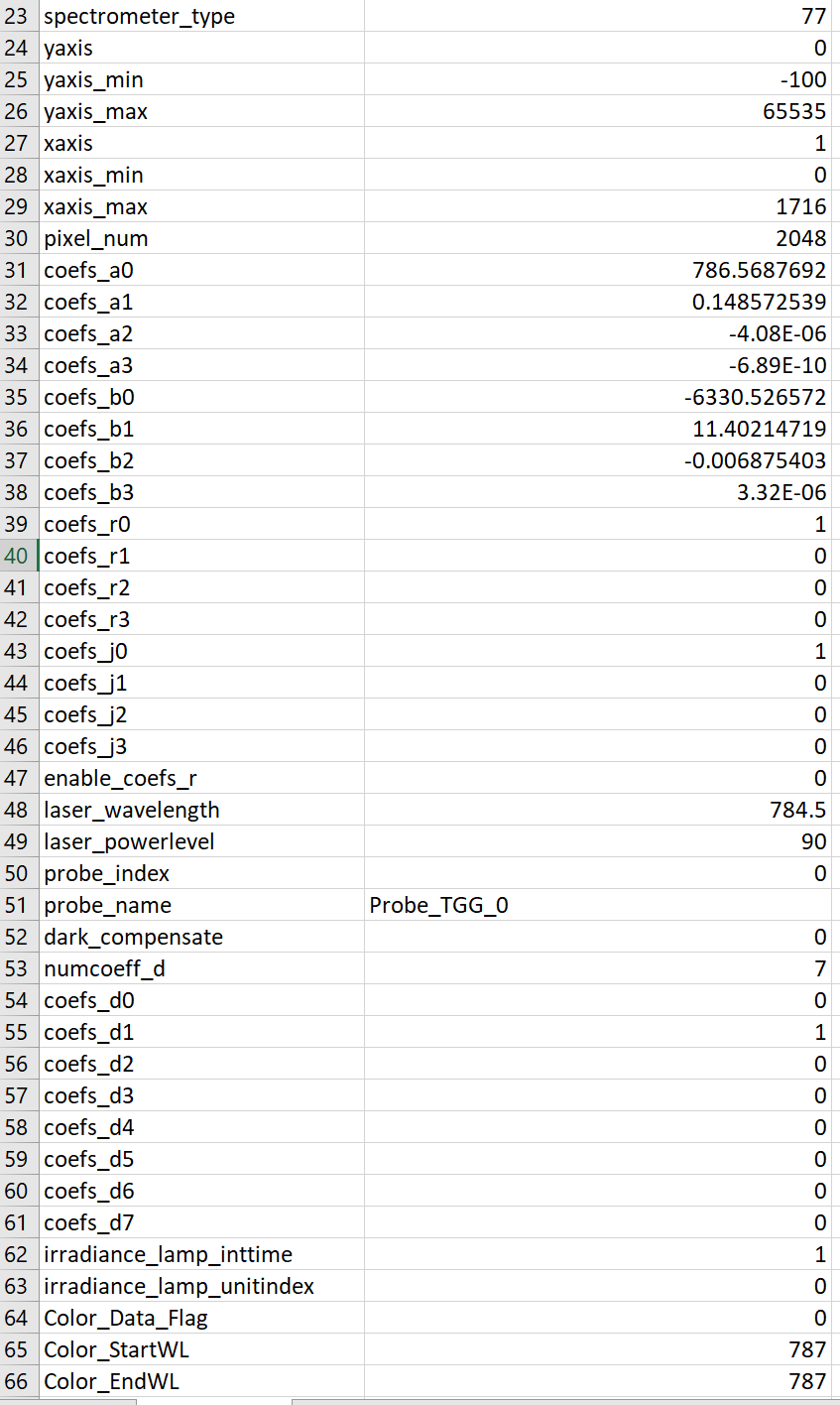
## Annex 1.

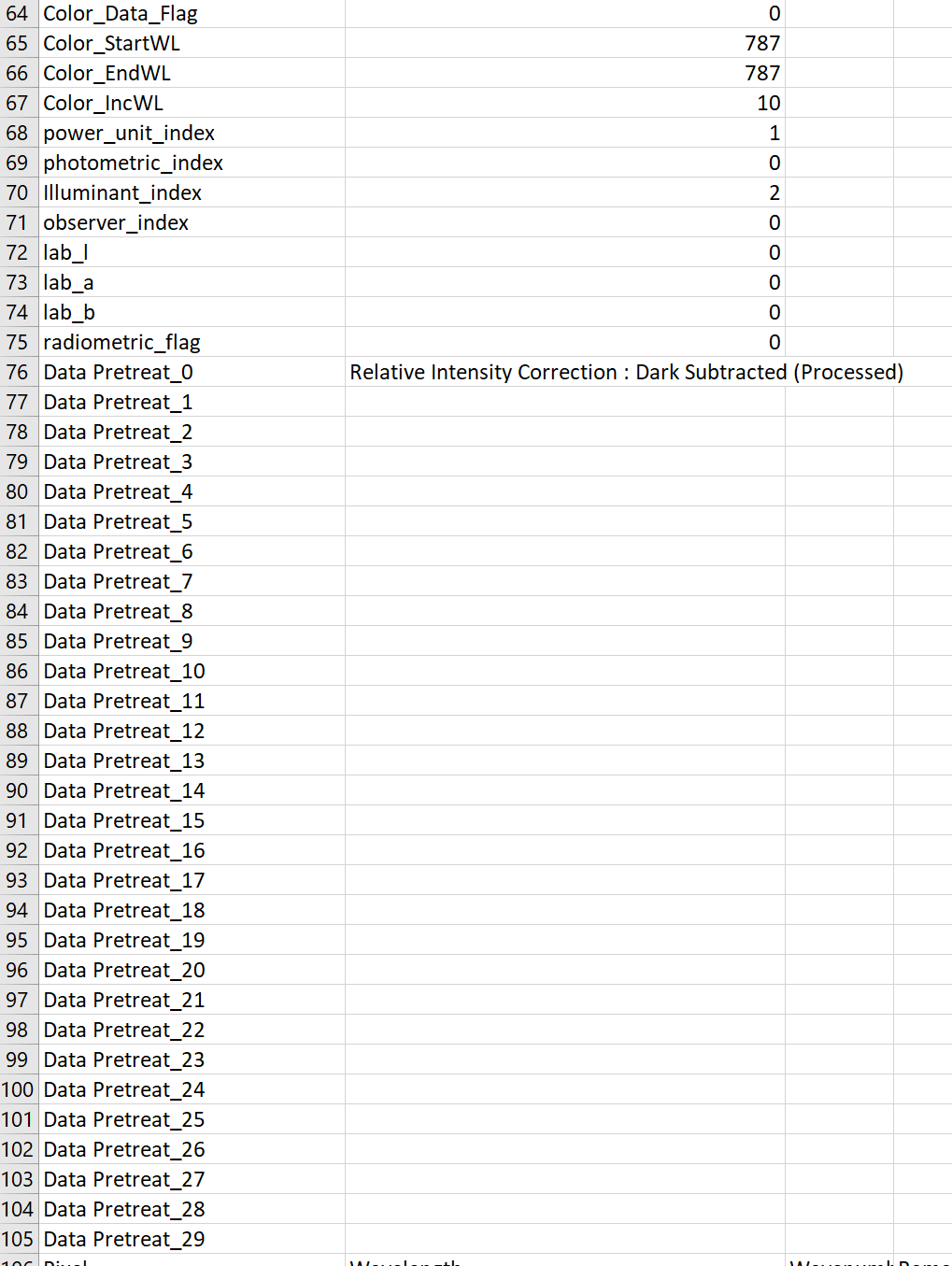
What are the actual files and what is required?

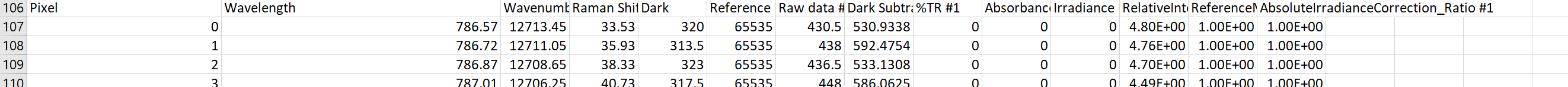
Analysing a -limited- number of files it seems that the pattern of the data files contains some logical similitudes but some strong differences. These will be analysed on this document and recommendations will be made.











Data to be available as txt files.

The section of metadata proposed can be:

* Section 1: manufacturer info and device tags (SN, model number…)
* Section 2: Sample and sampling information. Optical path
* Section 3: Spectrometer information and its corrections. Data points
* Section 4: Laser section
* Section 5: Algorithms available, used yes/no. here we need to include calibration and “spectra of the calibrations/corrections”
* Section 6: Data of the spectrum: nm (x-data, from spectrometer) / Raman shift (x-data, corrected)/ background (x-data, corrected) / intensity reading (a.u. units, incorrected, y-axis) / intensity (a.u. units, corrected)
  + How to have mapping and stage position???
* Section 7: Data transfer methods. chemometrics