CHARISMA

H2020 CHARISMA

Raquel Portela

CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

CHARISMA receives funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 952921

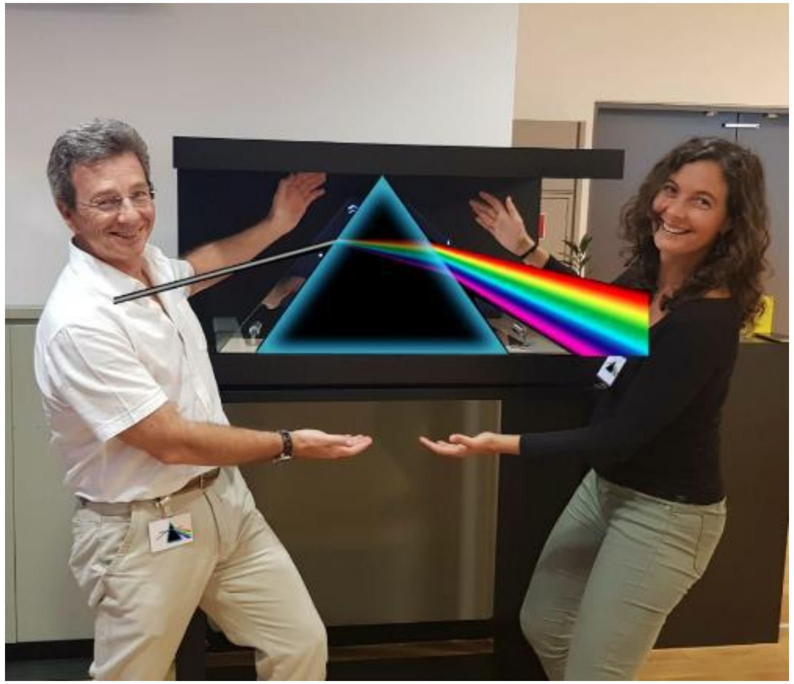
What is CHARISMA

Characterization and
HARmonization for
Industrial
Standardization and advanced
MAterials

Fact sheet:

- Call: H2020-NMBP-TO-IND-2020
- Topic: NMBP-35-2020
  Towards harmonised characterisation protocols in NMBP (RIA)
- Grant agreement: 952921
- Start date: 01.11.2020 (we‘re at M16)
- Duration: 48 months
- EU grant: 5 M€
- Beneficiaries: 14 (9 countries)
- Coordinator: CSIC: Miguel A. Bañares and Raquel Portela

Raman



CHARISMA objective

Harmonize Raman characterization, developing:

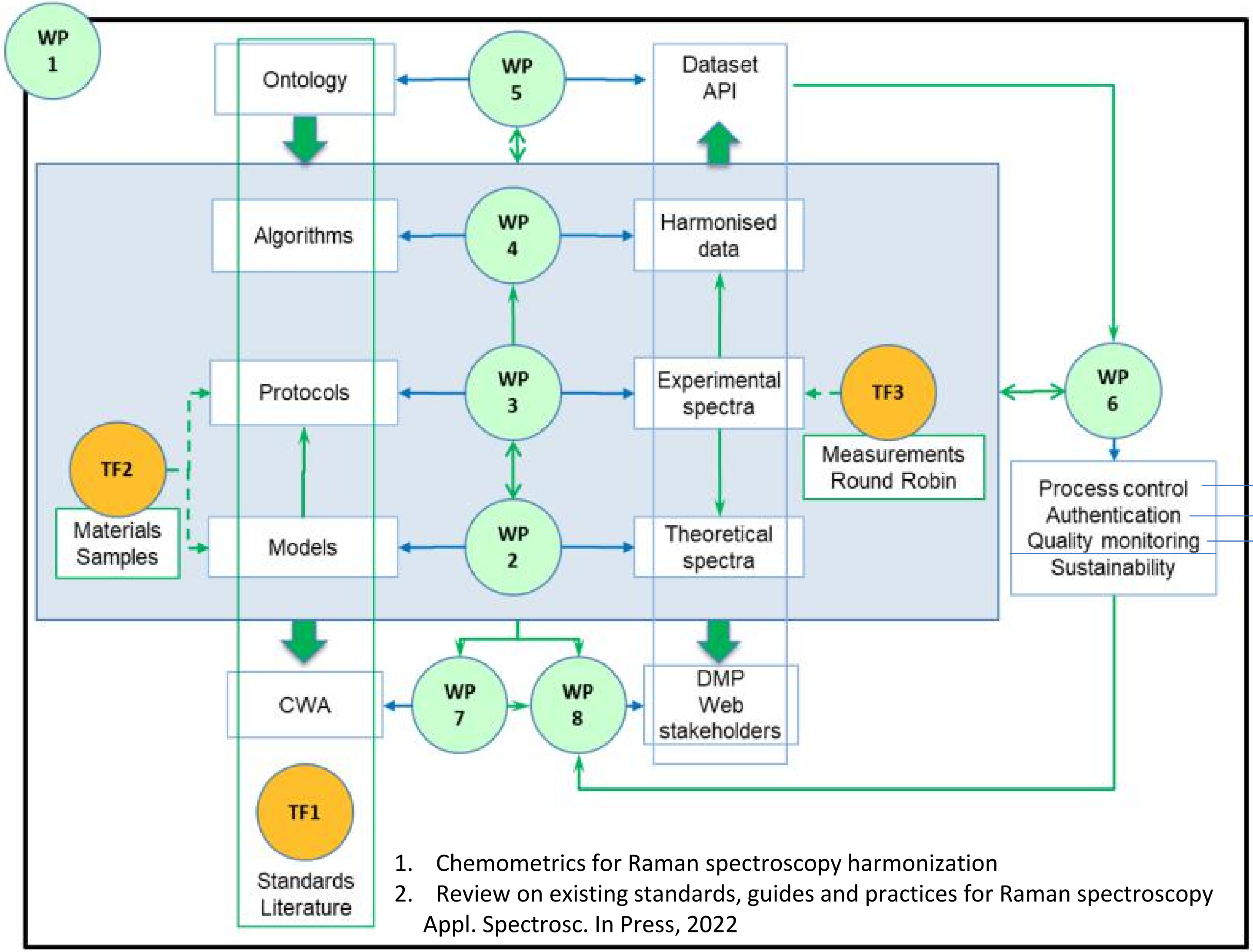
- **Algorithms** and **protocols** for spectra acquisition and data analysis
- **Reference master samples** and **theoretical spectra** for calibration and quantification
- **Hardware** for in-line, on-site measurement
- **Correlations** between Raman data and process or product descriptors
- **FAIR Raman data repository** to enable end users to share digital spectral data

CHARISMA impact

- **Foster the industrial implementation of Raman spectroscopy** as real-time, in-line and distributed monitoring and control/decision tool, **improving the business of existing products/processes** due to improved product quality and trust, waste minimization and time and energy saving.
- Raman harmonisation will also enable the fast development of **new business** based on Raman-active nanomaterials that face different societal challenges related to energy, security, or safety.

CHARISMA

CHARISMA Scheme



1. Chemometrics for Raman spectroscopy harmonization
2. Review on existing standards, guides and practices for Raman spectroscopy
   Appl. Spectrosc. In Press, 2022

- **Lack of interoperability**
- **Increasingly softer hardware**
- **Internet of things**

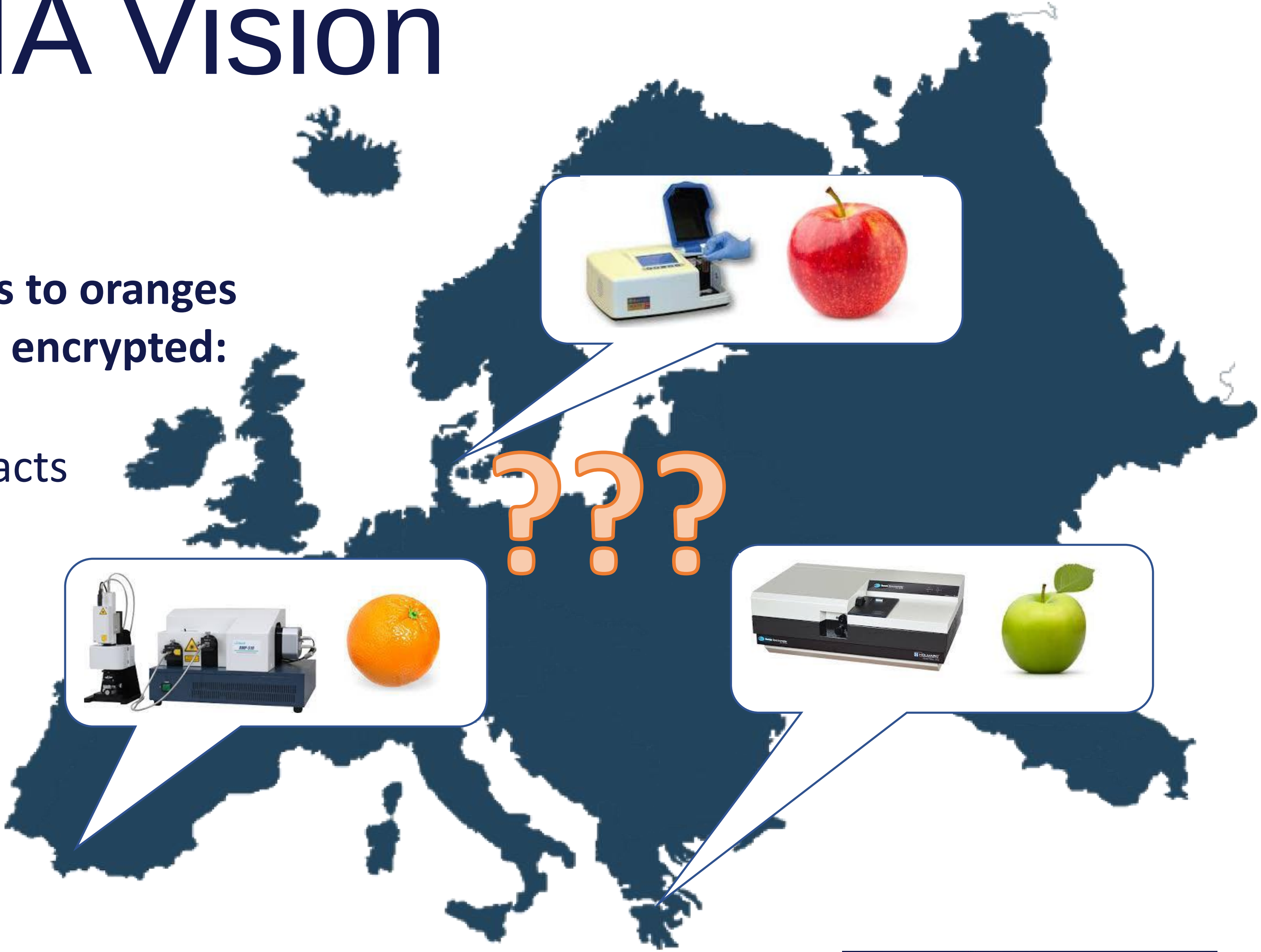
CHARISMA receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No **952921**

CHARISMA Vision

Raman in 2021

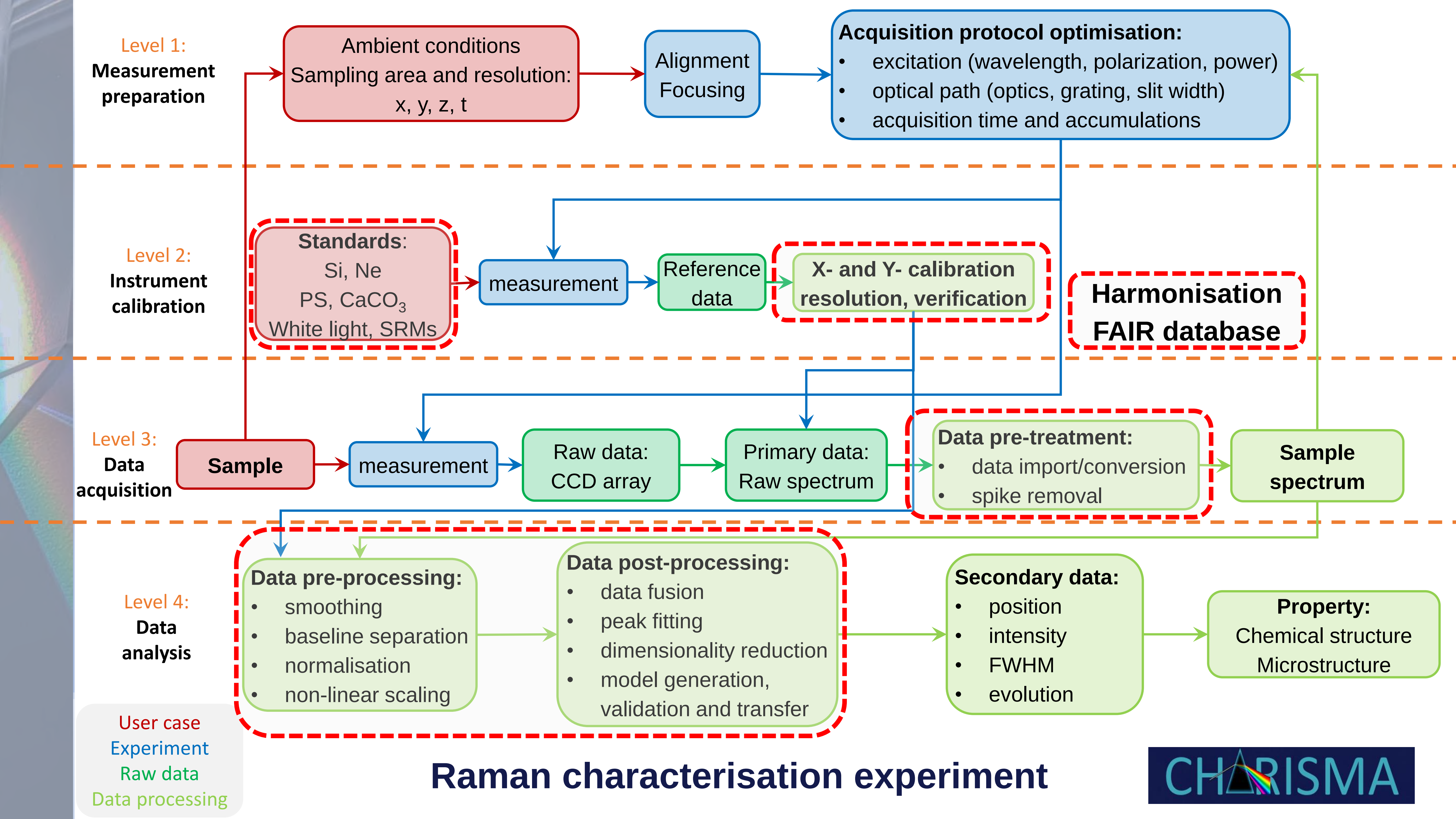
- We are **comparing apples to oranges**
- **Material fingerprints are encrypted:**

machine-
location- } specific artefacts
operator-



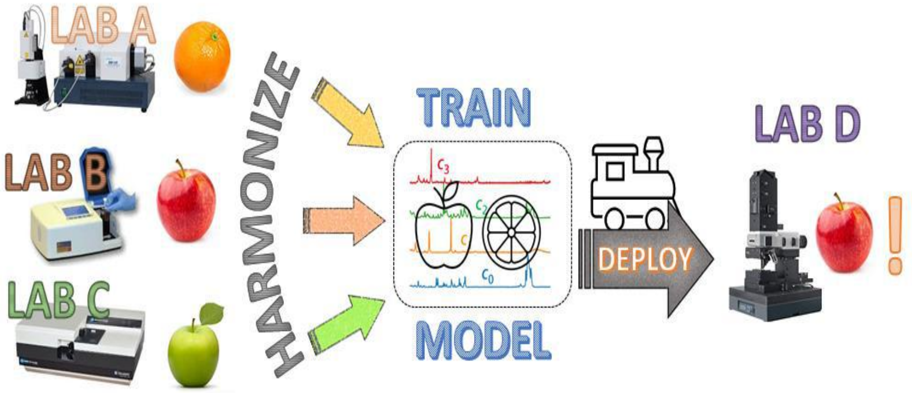
CHARISMA receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No **952921**

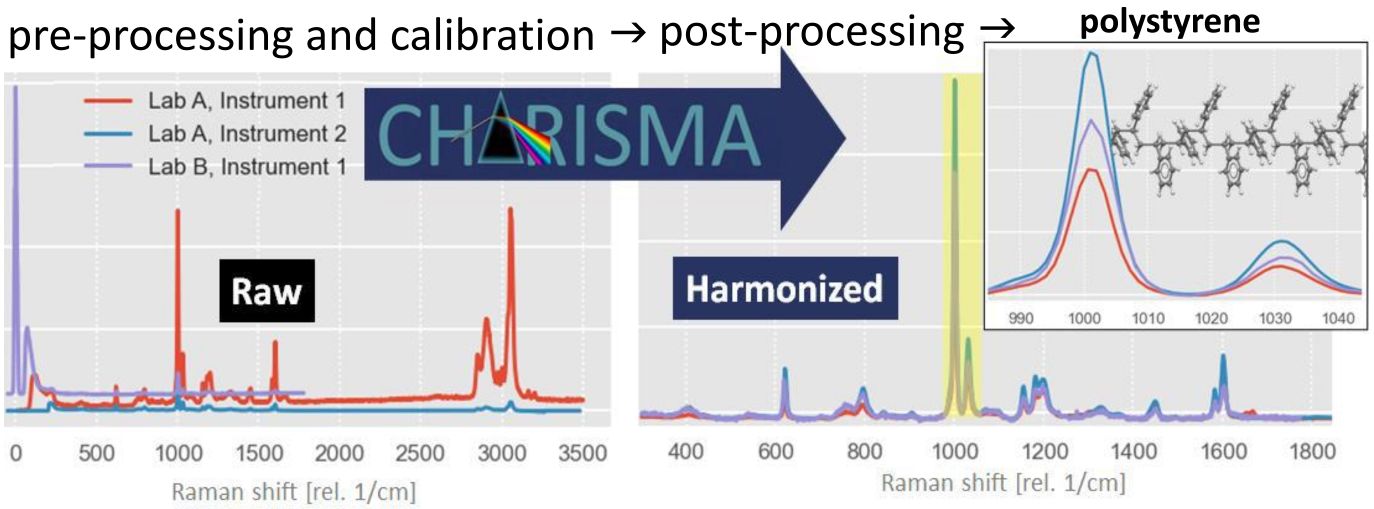
CHARISMA


Level 1:
Measurement preparation
Ambient conditions
Sampling area and resolution:
x, y, z, t
Alignment
Focusing
Acquisition protocol optimisation:
• excitation (wavelength, polarization, power)
• optical path (optics, grating, slit width)
• acquisition time and accumulations
Level 2:
Instrument calibration
Standards:
Si, Ne
PS, $CaCO_3$
White light, SRMs
measurement
Reference data
X- and Y- calibration
resolution, verification
Harmonisation
FAIR database
Level 3:
Data acquisition
Sample
measurement
Raw data:
CCD array
Primary data:
Raw spectrum
Data pre-treatment:
• data import/conversion
• spike removal
Sample spectrum
Level 4:
Data analysis
Data pre-processing:
• smoothing
• baseline separation
• normalisation
• non-linear scaling
Data post-processing:
• data fusion
• peak fitting
• dimensionality reduction
• model generation, validation and transfer
Secondary data:
• position
• intensity
• FWHM
• evolution
Property:
Chemical structure
Microstructure
User case
Experiment
Raw data
Data processing
Raman characterisation experiment
CHARISMA

Harmonization and database

- Development and production of universal, robust and readily available **calibration standards**
- Standardization of measurement **procedures**
- Easy data conversion to a **universal and open format**
- A **database** enabling exchange, collaboration and training of AI **prediction models**





Use case for the CHARISMA database

Data from different locations and instruments can be combined to generate **models** that predict material **properties**

Models can be deployed and applied anywhere using appropriate **transfer algorithms**

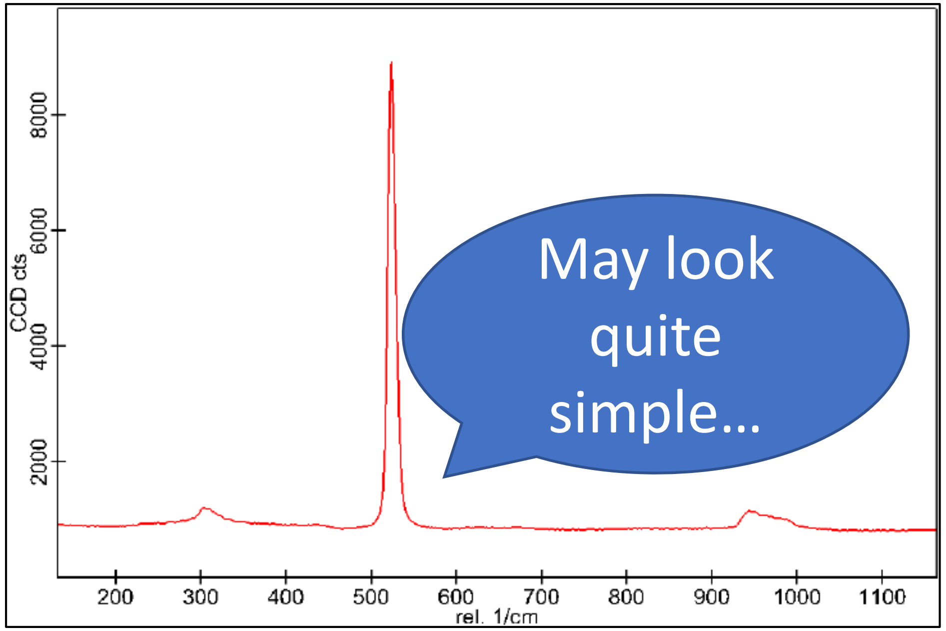
CHARISMA receives funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No **952921**

CHARISMA

Structure for harmonised Raman data:

Raman CHADA

Universally describe a Raman spectrum



k / cm^{-1}	λ / nm	CCD cts.
...	...	...
475.3	815.4	859
477.3	815.6	859
479.2	815.7	855
481.1	815.8	863
483.1	815.9	866
485.0	816.1	872
486.9	816.2	871
488.9	816.3	878
490.8	816.5	887
[illegible]	816.6	887
494.7	816.7	893
496.6	816.8	898
[illegible]	[illegible]	913
500.5	817.1	912
502.4	817.2	937
504.3	817.4	951
506.3	817.5	990
508.2	817.6	1046
510.1	817.7	1106
512.0	817.9	1253
514.0	818.0	1495
515.9	818.1	2115
517.8	818.3	3522
519.7	818.4	6044
521.7	818.5	8082
523.6	818.6	8927
525.5	818.8	8323
527.4	818.9	6739
529.3	819.0	5311
531.3	819.2	3876
533.2	819.3	2735
535.1	819.4	1910
537.0	819.5	1488
538.9	819.7	1259
540.8	819.8	1135
542.8	819.9	1075

Spectrum data

- **Raman shift range?**
- **Raman shift sampling?**
- **How many digits?**
- **Data format ?**
- **...**

General:

- System ID: 130-1200-830
- Time: 4:47:58 PM
- Date: Monday, November 30, 2020
- User Name: Mess_PA
- Sample Name: Si [100]
- Configuration: Raman 785

Spectrometer

- Excitation Wavelength [nm]: 785.000
- Grating: G1: 600 g/mm BLZ=750nm
- Center Wavelength [nm]: 847.998
- Spectral Center [rel. 1/cm]: 946.372

CCD Detector

- Width [Pixels]: 1024
- Height [Pixels]: 128
- Temperature [°C]: -59
- Integration Time [s]: 5.01222
- [illegible] 14907
- [illegible]
- [illegible] [8.25]
- [illegible] Shift Speed [MHz]: 0.1
- Preamplifier Gain: 1
- ReadMode: Full Vertical Binning
- Number Of Accumulations: 1
- Integration Time [s]: 5.01222

Optics

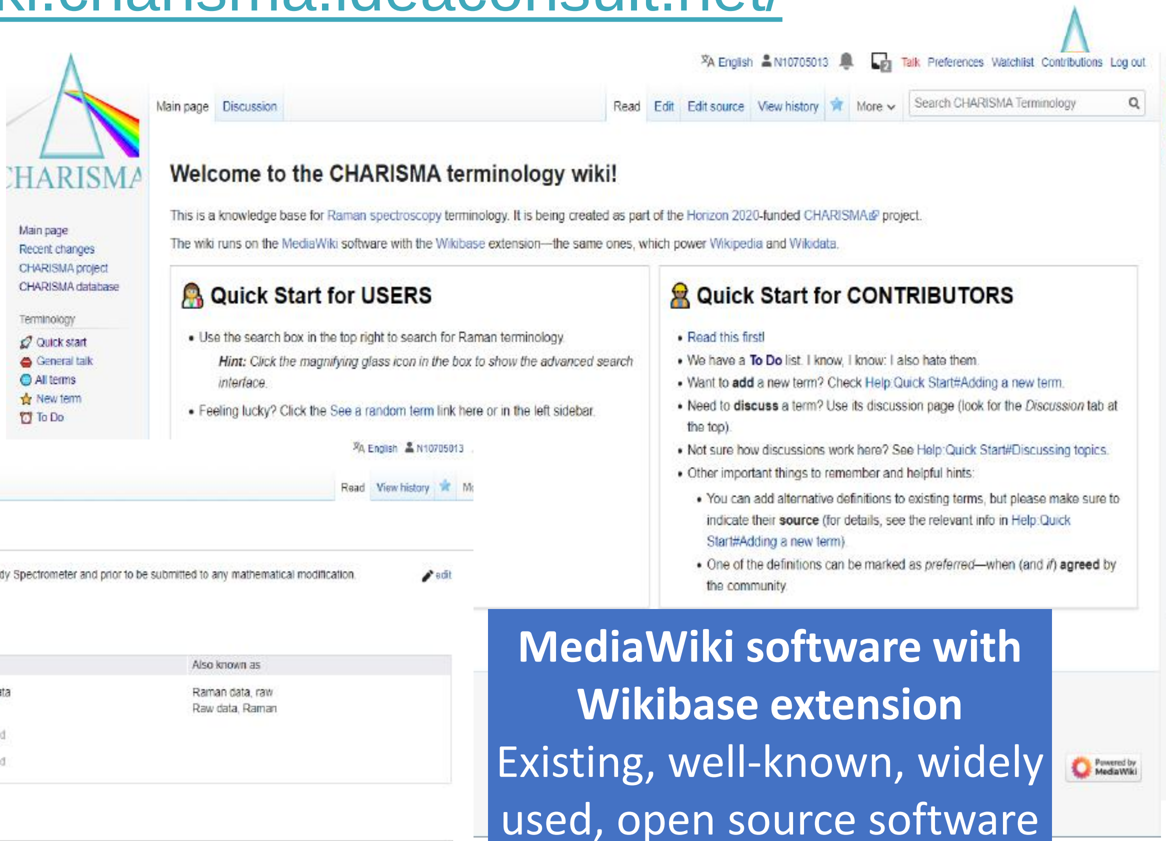
- Objective Name: Zeiss LD EC 50x / 0.55
- Objective Magnification: 50.0

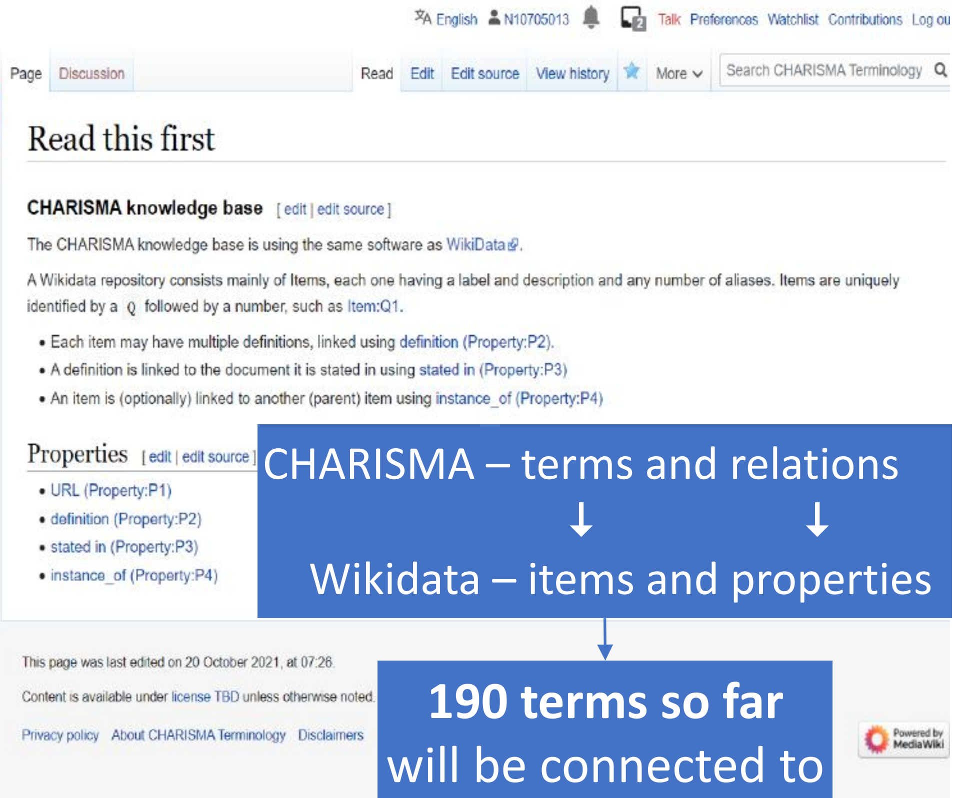
Meta data

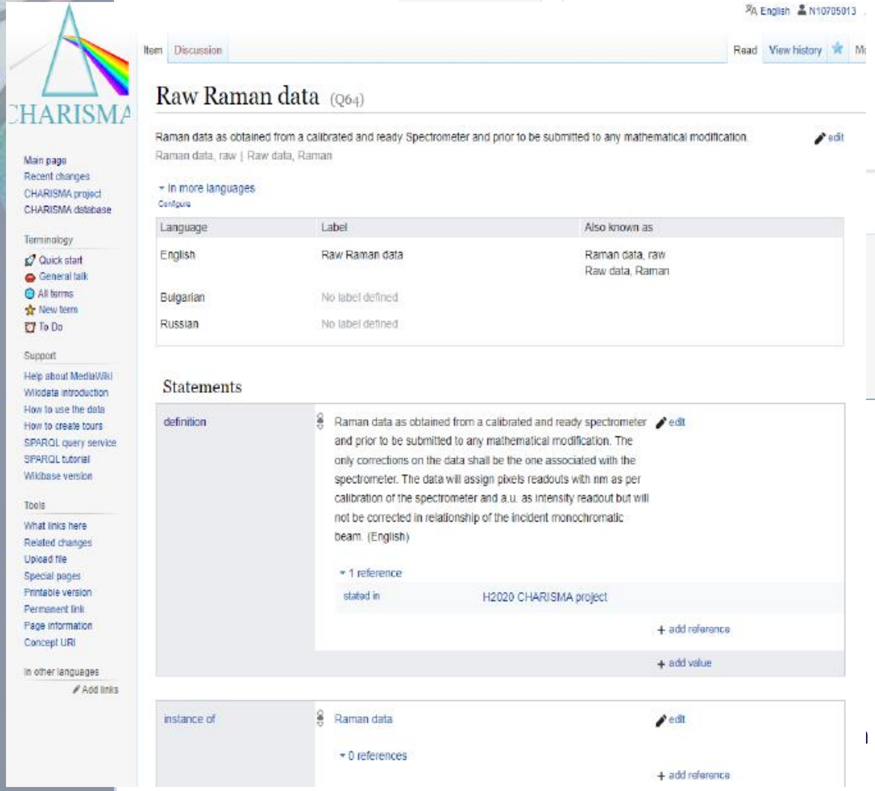
- **What does the system offer?**
- **What needs to be stored?**
- **Terminology ?**
- **...**

Wikidata for terminology harmonization

https://wiki.charisma.ideaconsult.net/







CHARISMA – terms and relations

↓ ↓

Wikidata – items and properties

↓

190 terms so far
will be connected to spectra metadata

MediaWiki software with Wikibase extension
Existing, well-known, widely used, open source software

- **User friendly UI to define structured knowledge and open to public**
- **SPARQL endpoint for programmers and integration**

Union's Horizon 2020 research and innovation programme under grant agreement No **952921**

CHARISMA

CHARISMA

RamanChada

Cheat Sheet

Loading & displaying data

`R = RamanChada('C:\file.spc')`

Spectrum object info: **R**

RamanChada with 3526 points generated Wed Sep 15 …

Get metadata as dict: `R.meta`

Add metadata (dict):
`R.add_metadata({'Power[mW]':5})`

Get processing log as list:
`R.log` *or* `R.show_log()`

Undo last processing step: `R.rewind(-1)`

Revert to original data: `R.rewind(0)`

Save .cha file with all changes:
`R.commit('commit message')`

Load .cha file:
`R = RamanChada('C:\file.cha')`

Plot spectrum: `R.plot()`

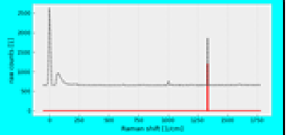
Get the raw data from the same file:

```
S = RamanChada(
    R.file_path, raw=True)
```

Cosmic rays

Fit x ray model
`R.fit_xrays()`

Plot x ray model
`R.plot_xrays()`



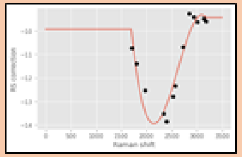
Remove x rays: `R.remove_xrays()`

Calibration

Interpolate to x axis of reference spectrum:
`R.interpolate_x(reference_spectrum)`

Calibrate x with existing RamanCalibration `cal`:
`R.calibrate(cal)`

Show x calibration curve:
`cal.show()`



Get calibration time / date:
`cal.time`

Generate x calibration using reference spec:
`cal = R.make_x_calibration(ref)`

…and consider only peaks in interval:

```
cal = R.make_x_calibration(
    ref,1700,3200)
```

Generate x calibration using peak positions list:

```
cal = R.make_x_calibration(
    [202.12,451.76,…,1809.28])
```

Calibrate y with existing RamanCalibration:
`R.calibrate_y(y_cal)`

Generate y calibration using reference spectrum:
`y_cal = R.make_y_calibration(ref)`

Save *RamanCalibration* to disk:
`cal.save('C:\cal_filename.chacal')`

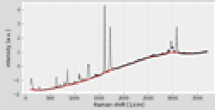
Load *RamanCalibration* from disk:

```
cal = read_x_calibration(
    'C:\cal_filename.chacal')
```

Baseline separation

Fit baseline using SNIP method
`R.fit_baseline(method='snip')`

Plot baseline model
`R.plot_baseline()`



Fit and remove baseline
`R.remove_baseline(method='snip')`

Pre-processing

Smooth spectrum using Savitzky-Golay filter:
`R.smooth('sg',window=11,order=3)`

Normalize spectrum using vector norm:
`R.normalize('vector')`

Area normalization using only an interval:
`R.normalize('area',500,1250)`

Crop spectrum on the x axis to 500-1.250 cm^{-1}:
`R.x_crop(500,1250)`

Add spectrum S to R: `R.math(S,'+')`

Peaks search & fitting

Find peaks with prominence>0.2 without fitting:
`R.peaks(prominence=0.2,fit=False)`

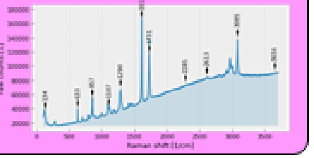
DataFrame with detected/fitted peaks: `R.bands`

Fit peaks within 2xFWHM w/ Voigt profile & plot:

```
R.peaks(fitmethod='voigt',
    interval_width=2,show=True)
```

Find peaks using wavelets, fit & sort by position:
`R.peaks(cwt=True,sort_by='position')`

Plot spectrum with peak positions
`R.show_bands()`



Batch processing

List of RamanChada objects from path list `files`:
`SL = [RamanChada(f) for f in files]`

Normalize all spectra by standard normal variate:
`[s.normalize() for s in SL]`

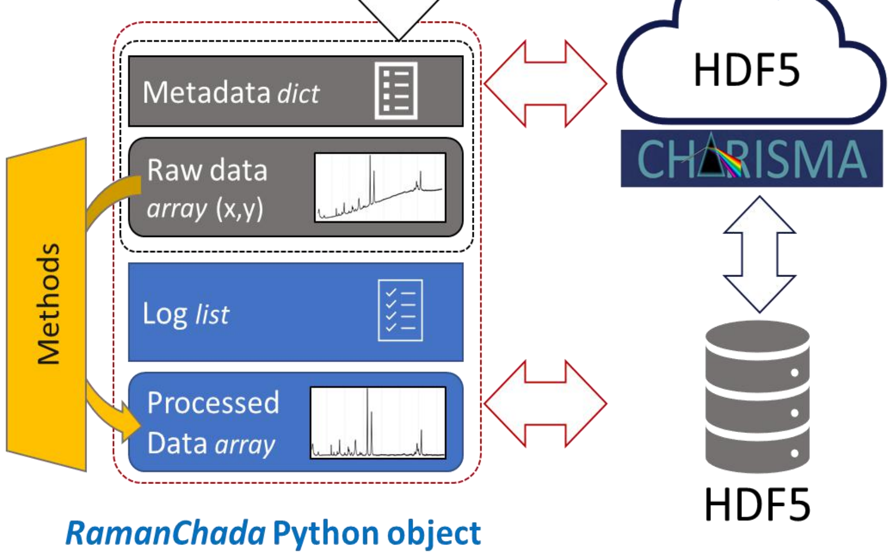
Apply Wiener filter to all and then save to disk:

```
[s.smooth('wiener',7) for s in SL]
[s.commit('smoothed') for s in SL]
```

spectra reading/writing and processing

Original file formats

- Text, CSV, JCAMP-DX
- Renishaw, BWTek, Horiba, Bruker, Thermo, …



CHARISMA receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No **952921**

CHARISMA

CHARISMA Database

https://search-dev.data.enanomapper.net/projects/charisma/

under development
dev-charisma@ideaconsult.net

Open source cloud-native HSDS for HDF5 storage

https://github.com/HDFGroup/hsds

- HSDS accessible via REST API (client **h5pyd** for hdf5 files)
- Python h5pyd library is a drop in replacement for **h5py**

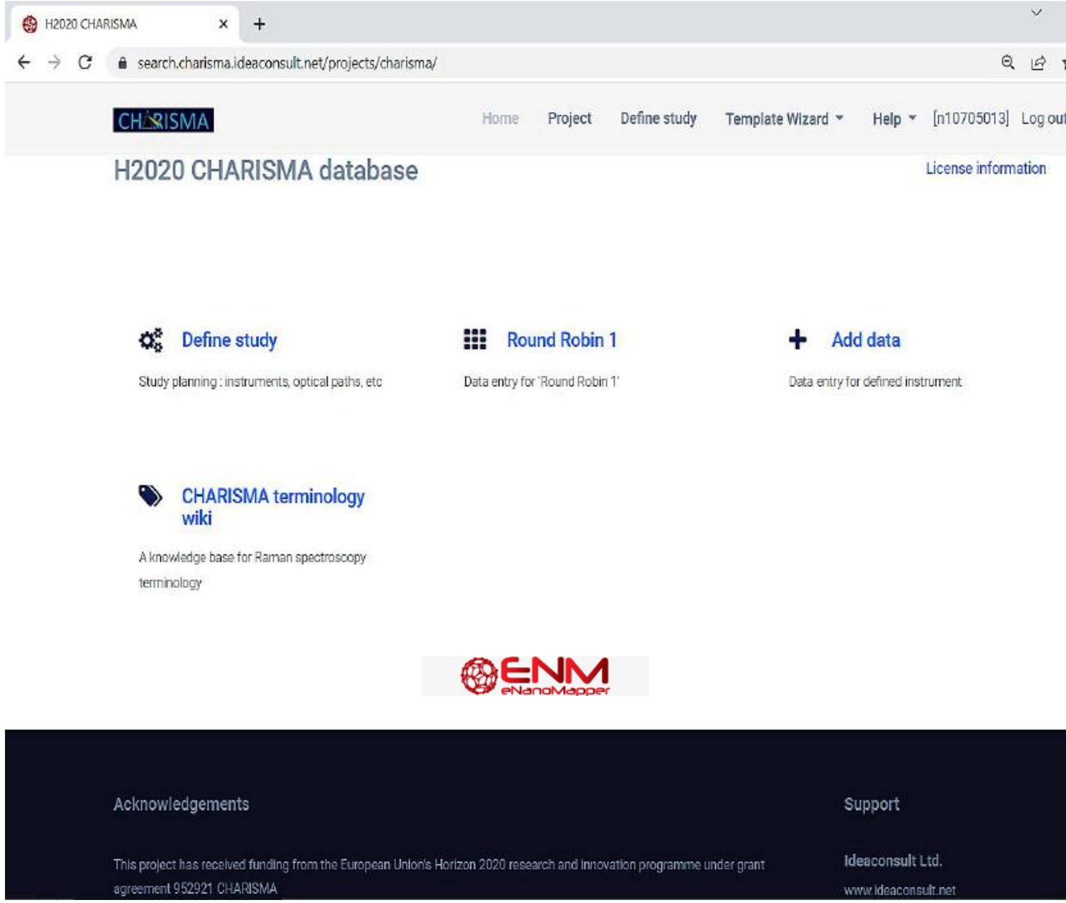
Flask API: wrapper for HSDS

api.charisma.ideaconsult.net

Batch upload: Python/Ploomber backend workflow

- converts native files to HDF5 and uploads to HSDS
- uses CHARISMA RamanChada Python package

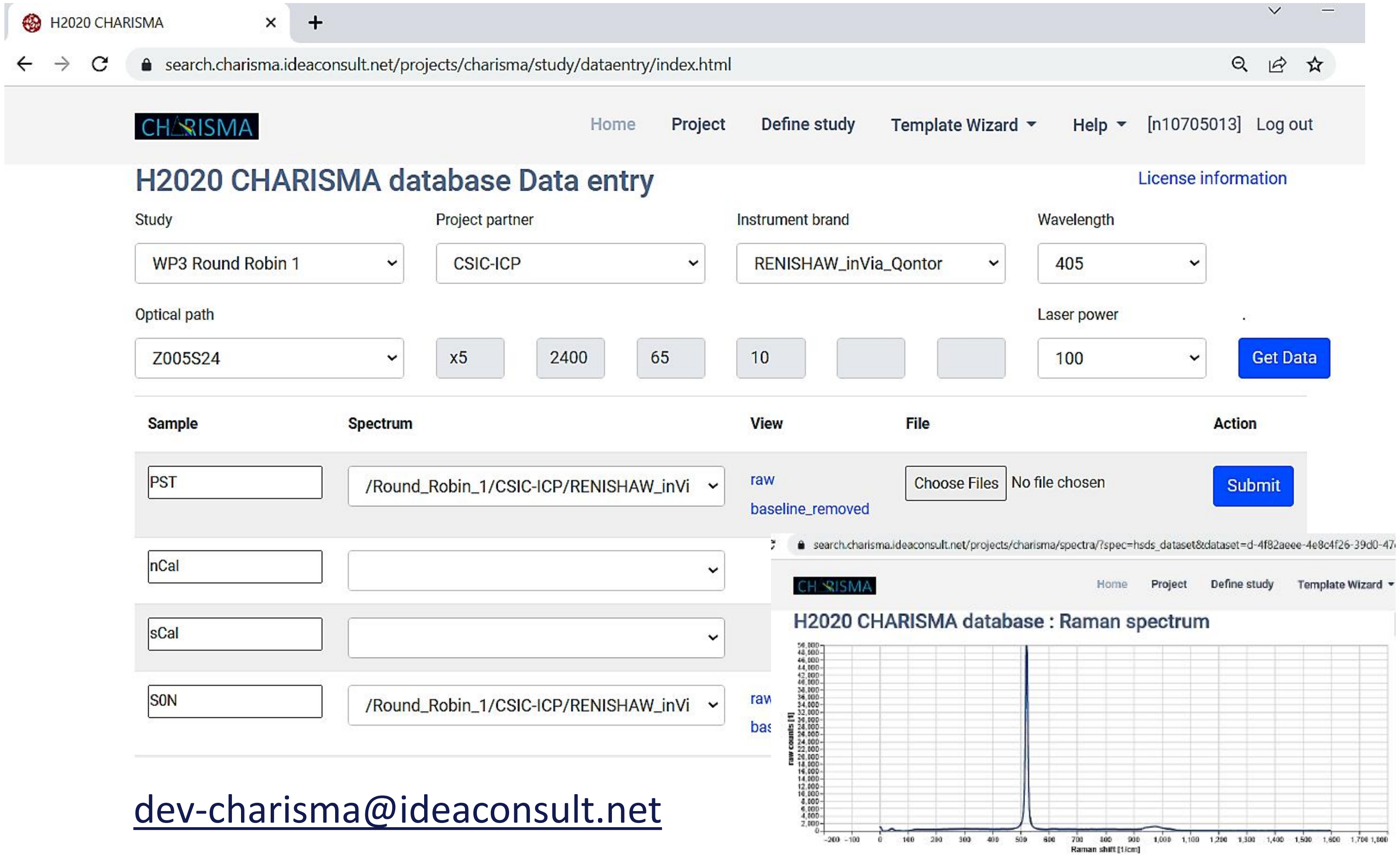
User interface: Jamstack (JavaScript, APIs and Markup)



RamanChada
HSDS
Flask
RESTful API
GET PUT POST DELETE
Ploomber
python
CHARISMA

CHARISMA Database UI & API

/api/dataset?domain=/Round_Robin_1/CSIC-ICP/RENISHAW_inVia_Qontor/405/



"subdomains": [
{
"name": "/Round_Robin_1/CSIC-ICP/RENISHAW_inVia_Qontor/405/211203.cha",
"annotation": [
{
"sample": "PST",
"instrument": "RENISHAW_inVia_Qontor",
"investigation": "Round_Robin_1",
"laser_power": "100",
"native_filename": "211203.wdf",
"optical_path": "Z005S24",
"provider": "CSIC-ICP",
"wavelength": "405"
}
],
"last_segment": "211203.cha",
"datasets": [
{
"key": "baseline_removed",
"uuid": "d-72aa93bb-74ed87f9-0fb6-bd7e54-0cadb7",
"name": "/baseline_removed",
"shape": [
2,
664
],
"size": 1328
},
{
"key": "raw",
"uuid": "d-72aa93bb-74ed87f9-3e7a-a12647-0096ef",
"name": "/raw",
"shape": [
2,
664
],
"size": 1328
}
]
},

dev-charisma@ideaconsult.net

CHARISMA database examples

RamanChada(

"/Round_Robin_1/CSIC-ICP/RENISHAW_inVia_Qontor/405/S0N_Qontor405_Z020H24_10_1x119800ms.cha")

```
domain = h5pyd.Folder("/Round_Robin_1/CSIC-ICP/RENISHAW_inVia_Qontor/405/")
print(domain)
domain._getSubdomains()
for d in domain._subdomains:
    print(d["name"])
```

[10] ✓ 0.2s

```
/Round_Robin_1/CSIC-ICP/RENISHAW_inVia_Qontor/405/
/Round_Robin_1/CSIC-ICP/RENISHAW_inVia_Qontor/405/211203.cha
/Round_Robin_1/CSIC-ICP/RENISHAW_inVia_Qontor/405/S0N_Qontor405_Z020H24_10_1x119800ms.cha
/Round_Robin_1/CSIC-ICP/RENISHAW_inVia_Qontor/405/S0N_Qontor405_Z020S24_100_1x10300ms.cha
/Round_Robin_1/CSIC-ICP/RENISHAW_inVia_Qontor/405/metadata.h5
```

```
with h5pyd.File("/Round_Robin_1/CSIC-ICP/RENISHAW_inVia_Qontor/405/metadata.h5") as metadata:
    for a in metadata.attrs:
        print(">{}={}".format(a,metadata.attrs[a]))
    ops = metadata["instrument"]["optical_paths"]
    for key in ops:
        print(key,ops[key])
        for op in ops[key]:
            for a in ops[key][op].attrs:
                print(">>{}\t{}\t{}={}".format(key,op,a,ops[key][op].attrs[a]))
```

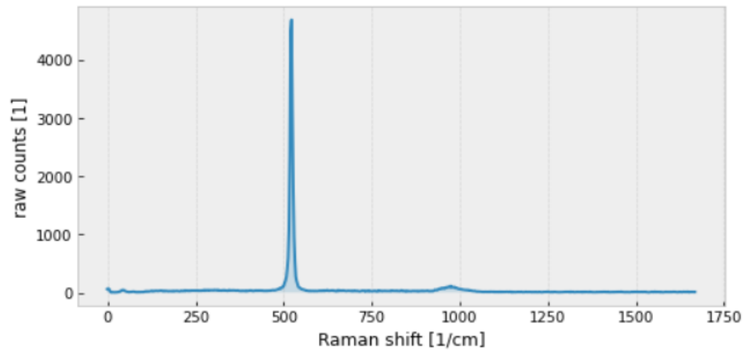
[22] ✓ 0.1s

```
>investigation=Round_Robin_1
>provider=CSIC-ICP
Z005S24 <HDF5 group "/instrument/optical_paths/Z005S24" (1 members)>
>>Z005S24       laser_power_1   settings=100
>>Z005S24       laser_power_1   power_mw=11
Z020S24 <HDF5 group "/instrument/optical_paths/Z020S24" (1 members)>
>>Z020S24       laser_power_1   settings=100
>>Z020S24       laser_power_1   power_mw=20
```

```
with h5pyd.Folder("/Round_Robin_1/CSIC-ICP/RENISHAW_inVia_Qontor/405/") as domain:
    #print(domain)
    domain._getSubdomains()
    for d in domain._subdomains:
        with h5pyd.File(d["name"]) as file:
            tmp=""
            for group in ["annotation_sample","annotation_study"]:
                for a in file[group].attrs:
                    tmp += "{} = {}; ".format(a,file[group].attrs[a])
        print(d["name"],"\n",tmp)
        R = RamanChada(d["name"],raw=True,is_h5pyd=True)
        R.plot()
```

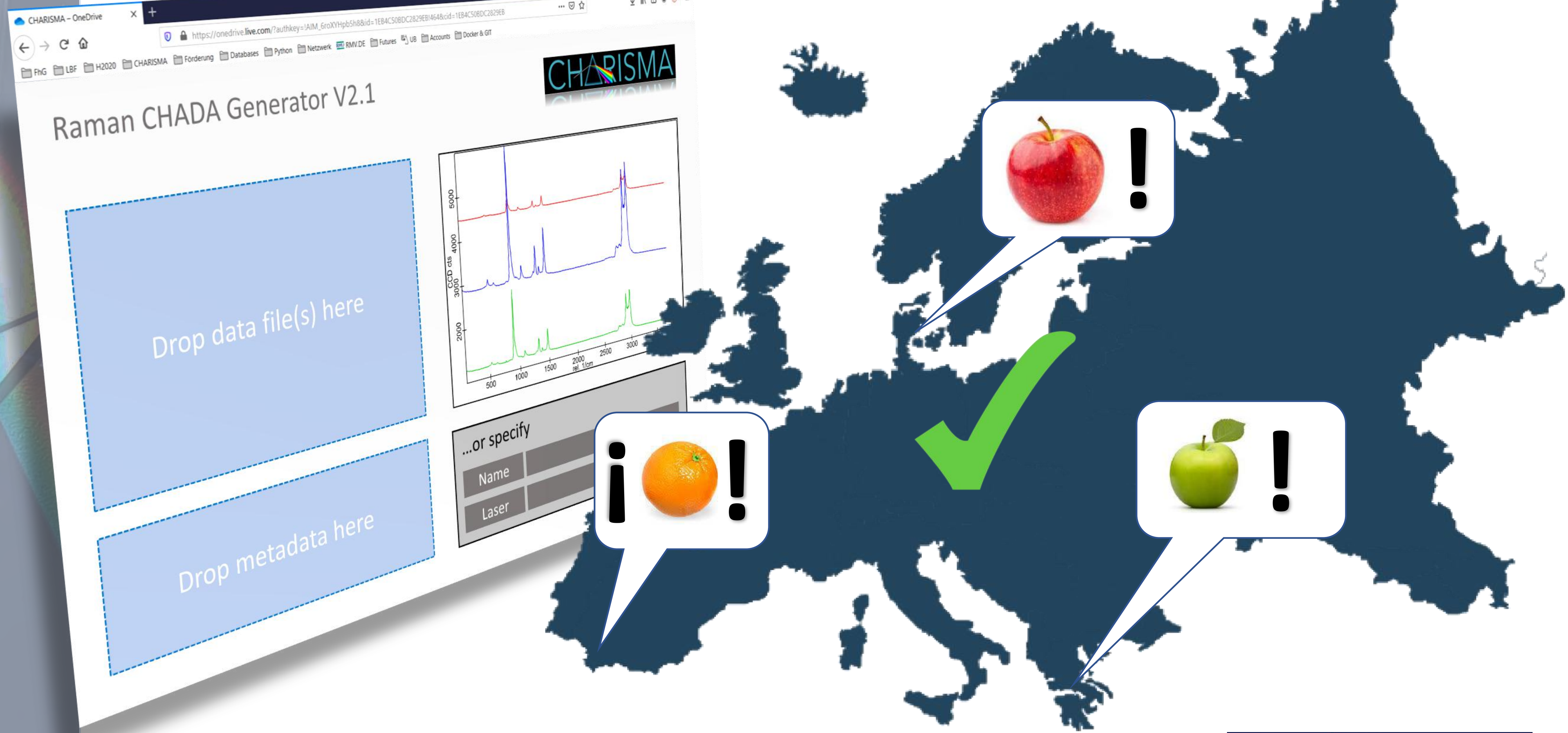
⊗ 1.6s

```
/Round_Robin_1/CSIC-ICP/RENISHAW_inVia_Qontor/405/S0N_Qontor405_Z020H24_10_1x119800ms.cha
 sample = PST; instrument = RENISHAW_inVia_Qontor; investigation = Round_Robin_1; laser_power = 100; native_filename =
S0N_Qontor405_Z020H24_10_1x119800ms.wdf; optical_path = Z005S24; provider = CSIC-ICP; wavelength = 405;
```



CHARISMA

Raman in 2024



CHARISMA receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No **952921**

Thank you!

CHARISMA

Presentation by Raquel Portela, co-coordinator

raquel.portela@csic.es, miguel.banares@csic.es

info@h2020charisma.eu

Follow us:

www.h2020charisma.eu

Twitter: @h2020charisma

LinkedIn: h2020-charisma

Questions?

CHARISMA receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952921