OPEN-SOURCE FOR RAMAN SPECTROSCOPY HARMONISATION

G. Georgiev[a,b], D. Lelinger[c], L. Iliev[a], V. Jeliazkov[a] , M. Banares[d], R. Portela[d], N. Jeliazkova[a]

[a] *Ideaconsult, 4 Angel Kanchev St, 1000 Sofia, Bulgaria*

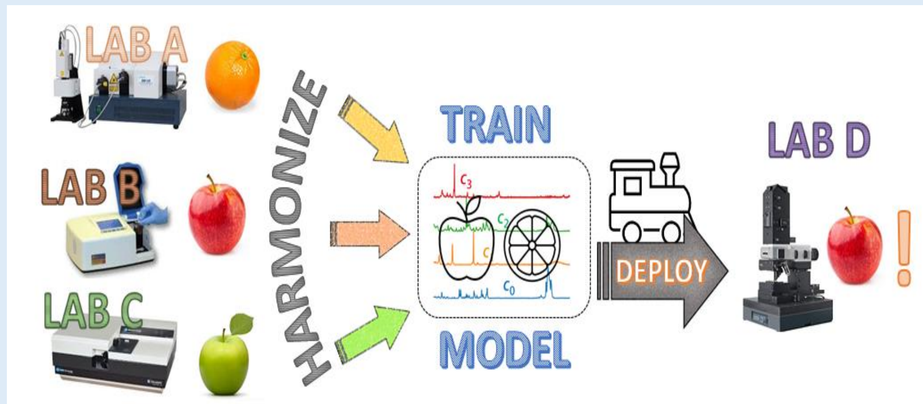
[b] *Institute for Nuclear Research and Nuclear Energy, Bulgarian Academy of Sciences, Tsarigradsko Shose 72, 1784 Sofia, Bulgaria*

[c] *Fraunhofer LBF, Plastics Division, Schloßgartenstraße 6, 64289 Darmstadt, Germany*

[d] *CSIC, Institute for Catalysis, Marie Curie 2, E-28049-Madrid, Spain*

BACKGROUND MOTIVATION AND OBJECTIVE

- Raman spectroscopy is becoming a key technology used in the research and development for characterization of materials.
- Ideally, spectra should be comparable and linked to specific material properties.
- In real life, Raman spectra differ between instruments and depend on the spectrometer, optical path, or sample environment, among others, and only a limited number of calibration standards for Raman spectroscopy are available to date [1]



The CHARISMA EU's Horizon 2020 project aims to

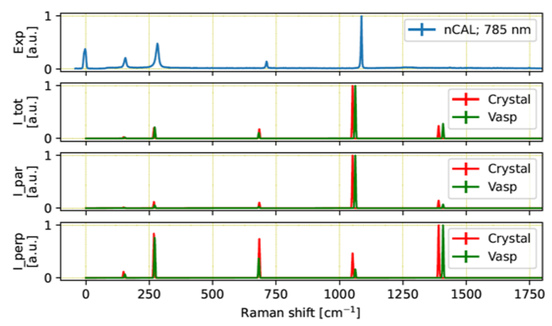
- harmonize and standardize characterization by Raman spectroscopy
- including hardware, measurement protocols, and in-silico methods
- enabling end users to share digital spectral data through a FAIR database across domains and across the entire life cycle of diverse products

METHODS

We present **ramanchada2**, an open-source, MIT-licensed, Python package
https://github.com/h2020charisma/ramanchada2
Ramanchada2 is published at the Python Package Index repository (PyPI).

HARMONISED SPECTRA

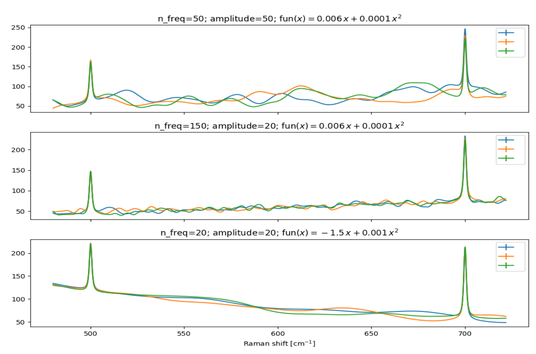
Raman spectra data input and output
- ✓ Multiple file formats supported
- ✓ Simulated data (VASP, Crystal)
- ✓ Harmonized HDF5 file format (output)



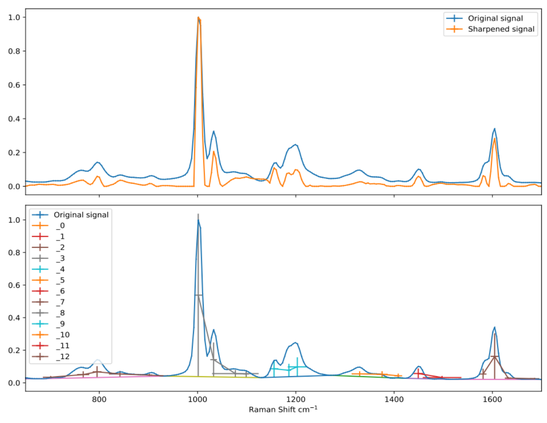
Spectra preprocessing
- ✓ Scale & normalize x-axis
- ✓ Baseline removal
- ✓ Resampling non-uniform DFT

```
spe_deltas = rc2.spectrum.from_delta_lines(deltas={800:1e3, 1800:1.5e3, 2500:1e3})
spe_initial = spe_deltas.resample_NUDFT_filter(xnew_bins=100).convolve(lineshape='voigt', sigma=.1, gamma=1)
spe_baseline = spe_initial.add_baseline(n_freq=30, amplitude=50, pedestal=10, func=lambda x: x*.05 - .00001*x**2)
spe_complete = spe_baseline.add_poisson_noise(scale=.5)
spe_scaled = spe_complete.scale_xaxis_fun(lambda x: x**2*.0001 + x*.22)
spe_nobaseline = spe_scaled.subtract_moving_minimum(6)
spe_resampled = spe_nobaseline.resample_NUDFT_filter(xnew_bins=100, x_range=(0, 2500))
spe_normalized = spe_resampled.normalize()
```

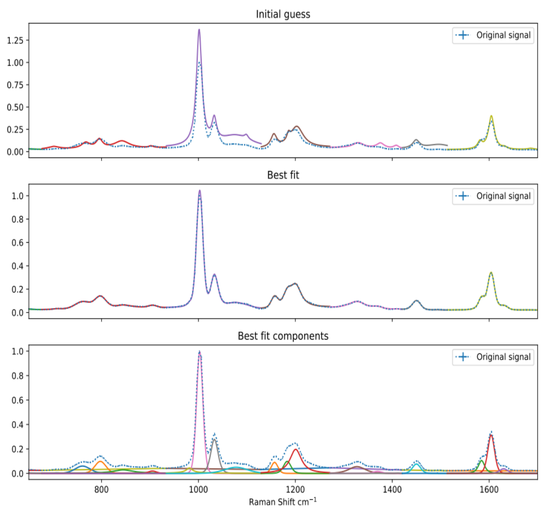
Synthetic spectra generation
- ✓ User defined peaks (position, intensities)
- ✓ Convolve with peak profiles (Voigt, Gaussian, Pearson IV, VII, Moffat, User defined function)
- ✓ Add artificial baseline, noise



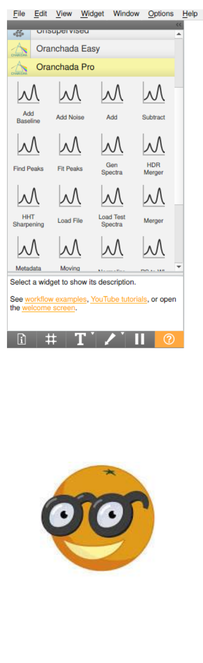
Find peak candidates
- ✓ Sharpening
- ✓ Bayesian Gaussian mixtures
- ✓ Peak finding
- ✓ Group overlapping peaks

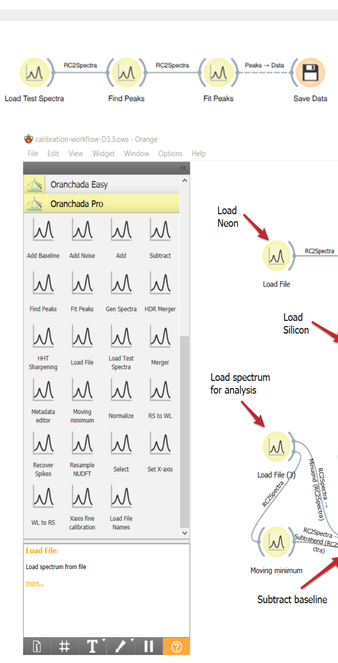


Peak fitting
- ✓ Composite model
- ✓ Fit on sections
- ✓ Models available: Gaussian, Voigt, Pearson, Moffat

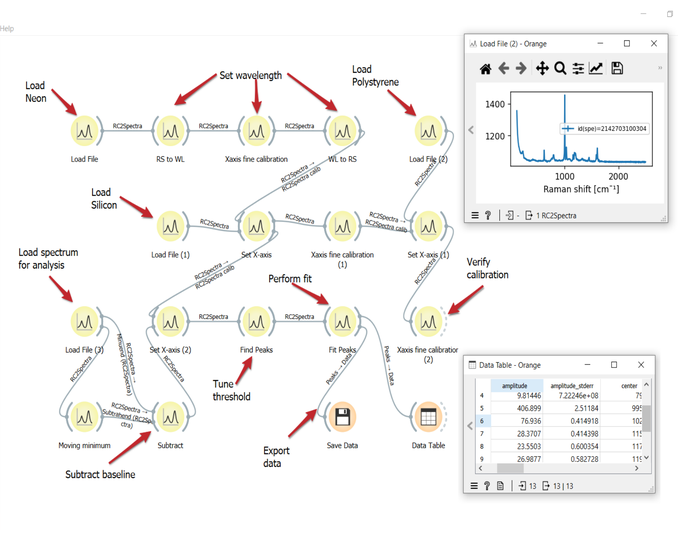


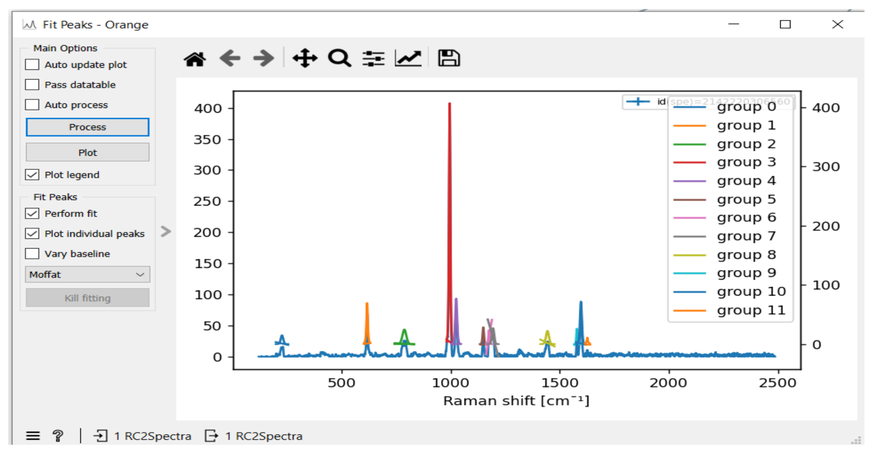
To facilitate Raman spectra analysis for the end users, we are developing **Oranchada**, an user friendly wrapper of all ramanchada2 functionality, which is available for installation as an add-on of the popular **Orange data mining software** [2]. The Oranchada package is also open source and MIT-licensed and actively developed at **https://github.com/h2020charisma/oranchada.**



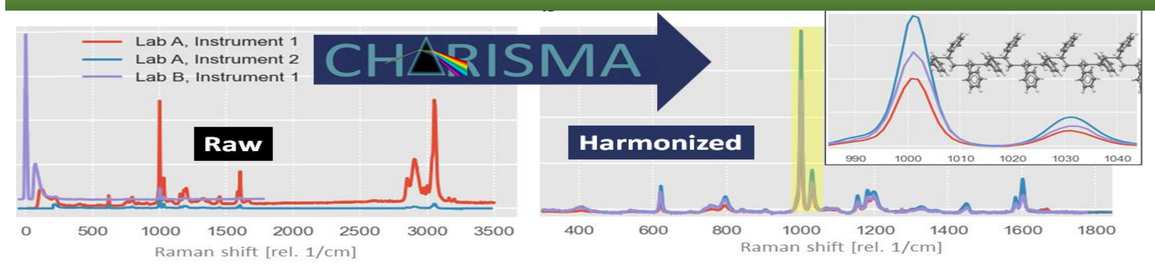


Calibration workflow



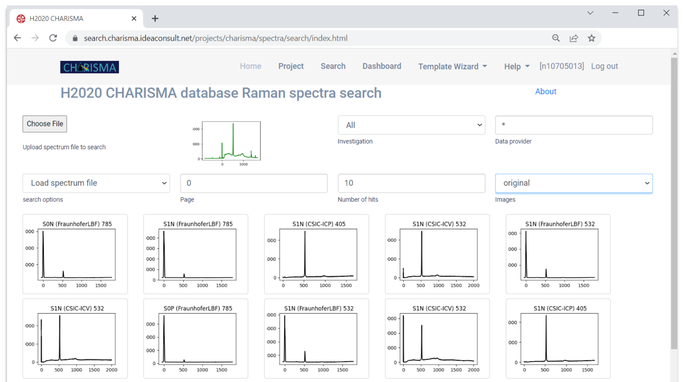


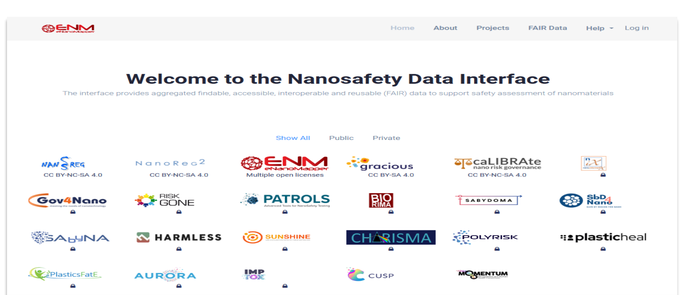
Results



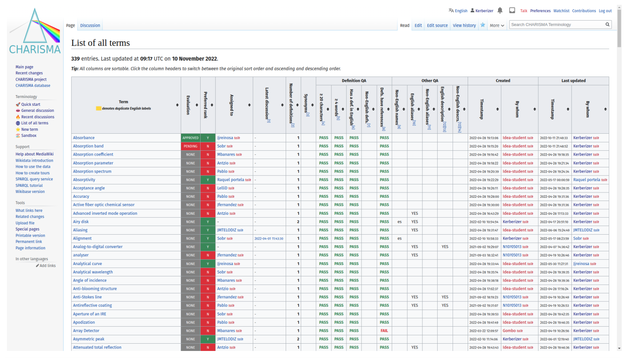
FAIR data (eNanoMapper compatible)

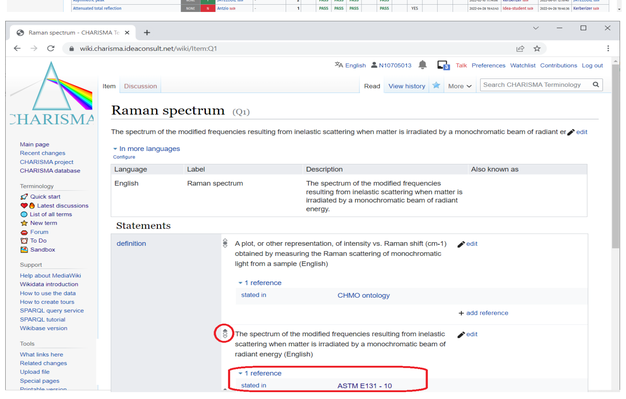
raw and harmonized spectra





Wikibase for terminology harmonisation





- ✓ Open source for Raman spectra harmonisation, calibration workflow
- ✓ Chemometric, e.g. qualitative analysis for chemical component mixtures
- ✓ ramanchada2 is used as a core dependency of a FAIR Raman database, implemented as Python API with a cloud backend for storing Raman spectra and metadata in HDF5 format. The metadata query and spectra search API is compatible with eNanoMapper database and NanoSafety Data Interface [4].
- ✓ For the harmonisation of Raman spectroscopy terminology, a collaboratively edited knowledge base is established at https://wiki.charisma.ideaconsult.net/. It is implemented with Wikibase, an extension for the MediaWiki software (https://www.mediawiki.org/), which is used by many popular wikis and, most notably, by Wikipedia.

1. A. Ntziouni et al., Applied Spectroscopy 76 (2022) 747-772 https://doi.org/10.1016/j.toxlet.2019.09.002
2. https://orangedatamining.com/
3. Kochev, N et al. *Nanomaterials*, *10*, **2020**, https://doi.org/10.3390/nano10101908
4. Jeliazkova, N et al. *Nat. Nanotechnol. 16, **2021**, 644–654* https://doi.org/10.1038/s41565-021-00911-6

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