

Task 4.3 Summary and Status

Martino Romaniello – ESO On behalf of Task 4.3





Task 4.1 Integration of astronomy VO data and services into the EOSC

Lead: Marco Molinaro (INAF)

Task 4.2 Implementation of FAIR principles for ESFRI data through the Virtual Observatory

Lead: Françoise Genova (CNRS-ObAS)

Task 4.3 Adding value to trusted content in astronomy archives

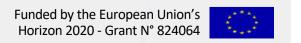
Co-leads: Mark Allen (CNRS-ObAS) & Martino Romaniello (ESO)



Task 4.3 Adding value to trusted content in astronomy archives

Next generation functionalities for creation and publication of **high-level, value-added data products** from ESFRIs

- Assessment and application of new techniques machine learning analytics. (connection to WP3)
 - Specific example applied to ESO archives data products: spectra, cubes, source catalogues
- Identification of stewardship best practices
 - Curation and publication of next-generation data products via ESFRI archives
 - Technical and human aspects





Major milestones

D4.5 Release of prototype machine learning enabled archive services providing value-added content to archives (Demonstration)	Month 30 July 2021
D4.8 Final analysis report on use of IVOA standards for FAIR ESFRI and community data and best stewardship practices for value-added data (Report)	Month 40 May 2022
(includes report of feedback on prototype services developed for D4.5)	

4



Deep Learning & Data Archives

- Scope: provide archive users with novel ways to identify data
 - Beyond traditional approach of specifying query parameters
 - Revolutionary extension of recent move from instrument to data keywords (e.g. exposure time to signa-to-noise)
- Target: the different data types in the ESO Archives
 - Spectra
 - Cubes
 - Source catalogues
- Let the data speak ... on a massive scale



Deep Learning on 1D spectra: HARPS

- Scope: Deep Learning analysis of the entire HARPS archive
 - High-resolution, high-stability spectrograph
 - Main science case: discovery and characterization of exoplanets
 - ~270,000 spectra, ~300,000 wavelengths channels each
 - 1D spectra, pipeline-reduced to high accuracy
- Different approaches
 - HITS: fully-connected autoencoder with 2 latent dimensions, down sampled spectral resolution
 - Agile architecture for speed and interactivity
 - ESO: combination of convolutional and fully-connected layers with between 4 to 8192 latent dimensions, full spectral resolution
 - Find minimal representation which preserves all the relevant information





So now, how to turn this into user services?

- These techniques have the potential of being a game-changer, BUT:
 - The results have to be meaningful
 - Interaction with the results has to be user-friendly

In practice:

- What makes spectra similar?
 - What are the uses and limitations of a similarity service?
- Can spectra be tagged with some physical properties of the celestial object?
 - If so, what are they?
 - What are the uses and limitations?
- What is the best way to present and interact with the results?