NEANIAS
Novel EOSC services for Emerging Atmosphere, Underwater and Space Challenges

Deliverable Report

Deliverable: D4.1 Space Research Services Report on Requirements, Specifications & Software Development Plan.

30/04/2020

NEANIAS is funded by European Union under Horizon 2020 research and innovation programme via grant agreement No. 863448

Document Info

Project Information			
Acronym	NEANIAS		
Name	Novel EOSC Services for Emerging Atmosphere, Underwater & Space Challenges		
Start Date	1 Nov 2019	End Date	31 Oct 2022
Program	H2020-EU.1.4.1.3. - Development, deployment and operation of ICT-based e-infrastructures		
Call ID	H2020-INFRAEOSC-2018-2020	Topic	H2020-INFRAEOSC-2019-1
Grant No	863448	Instrument	RIA
Document Information			
Deliverable No	D4.1		
Deliverable Title	Space Research Services Report on Requirements, Specifications & Software Development Plan.		
Due Date	30-Apr-2020	Delivery Date	07-May-2020
Lead Beneficiary	ALTEC (19)		
Beneficiaries (part.)	INAF (3); JACOBSUNI (13); UNIMIB (11); UOP (8); SZTAKI (4); MEEO (17)		
Editor(s)	Eugenio Topa (ALTEC); Eva Sciacca (INAF)		
Authors (s)	Eva Sciacca, Filomena Bufano, Eugenio Schisano (INAF); Rosario Messineo (ALTEC); Carlos Brandt (JACOBSUNI); Simone Mantovani (MEEO)		
Contributor (s)	Simone Riggi, Cristobal Bordiu, Ugo Becciani (INAF); Angelo Rossi (JACOBSUNI); Giuseppe Vizzari (UNIMIB); Mel Krokos (UOP); Robert Lovas (SZTAKI); Angelo Mulone (ALTEC)		
Reviewer(s)	Ricardo Vitorino (UBIWHERE)		
Workpackage No	WP4-Space		
Version	V1.0	Stage	Final
Version details	Revision: *2* . Last save: 2020-30-04 *, 07:57* Pages: *48* . Characters: *78.564*		
Distribution	Public	Type	Report
Keywords	EOSC; Services; Cloud Computing; Space Research;		

Change Record

Version	Date	Description	Editor
1.0	7/5/2020	Document version submitted to EC	Eva Sciacca

Disclaimer

NEANIAS is a Research and Innovation Action funded by the European Union under Horizon 2020 research and innovation programme, via grant agreement No. 863448.

NEANIAS is a project that comprehensively addresses the 'Prototyping New Innovative Services' challenge set out in the 'Roadmap for EOSC' foreseen actions. It drives the co-design, delivery, and integration into EOSC of innovative thematic services, derived from state-of-the-art research assets and practices in three major sectors: underwater research, atmospheric research and space research. In each sector it engages a diverse set of research and business groups, practices, and technologies and will not only address its community-specific needs but will also enable the transition of the respective community to the EOSC concept and Open Science principles. NEANIAS provides its communities with plentiful resource access, collaboration instruments, and interdisciplinary research mechanisms, which will amplify and broaden each community's research and knowledge generation activities. NEANIAS delivers a rich set of services, designed to be flexible and extensible, able to accommodate the needs of communities beyond their original definition and to adapt to neighbouring cases, fostering reproducibility and re-usability. NEANIAS identifies promising, cutting-edge business cases across several user communities and lays out several concrete exploitation opportunities.

This document has been produced receiving funding from the European Commission. The content of this document is a product of the NEANIAS project Consortium and it does not necessarily reflect the opinion of the European Commission. The editor, author, contributors and reviewers of this document have taken any available measure in order for its content to be accurate and lawful. However, neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated in the creation and publication of this document may be held responsible for any damage, financial or other loss or any other issue that may arise as a result of using the content of this document or any of the project outputs that this document may refer to.

The European Union (EU) was established in accordance with the Treaty on the European Union (Maastricht). There are currently twenty-seven (27) member states of the European Union. It is based on the European Communities and the member states' cooperation in the fields of Common Foreign and Security Policy and Justice and Home Affairs. The five main institutions of the European Union are the European Parliament, the Council of Ministers, the European Commission, the Court of Justice, and the Court of Auditors (http://europa.eu.int/).

Table of Contents

Tables of Figures & Tables

Document Figures

Document Tables

Abstract

The objective of deliverable D4.1 is to report on requirements, specifications and software development plan of the Space Research Services aimed at i) easing the management of astrophysics and planetary data through visualization (S1), ii) processing raw data and images toward generating multidimensional maps and mosaics (S2), and iii) automatize detection of structures within the maps (S3). This document reports on Space end-user communities, involved datasets and products and the current status of employed software and services. Then it presents the mapping of requirements to technical specifications and finally, the foreseen software development plan.

1. Introduction

1.1. Context

The NEANIAS WP4 "Space Research Services" is focused on the co-design of three innovative services in the Space environment for user-communities for Research and Development (R&D). The foreseen services will deliver a springboard of tools to enrich the workflows of a wide range of targeted users from academic/research institutions to industrial stakeholders, e.g. aerospace engineering and related technology companies, but also national space agencies and public outreach bodies, e.g. space museums and planetariums.
The first service, S1, will provide the required missing functionalities that enable efficient and scalable visual discovery, exposed through advanced interaction paradigms also exploiting virtual/augmented reality. The second service, S2, will deliver multi-dimensional space maps through novel mosaicking techniques (i.e. stitching of multidimensional images/maps with overlapping fields of view) to a variety of prospective users/customers (e.g., mining & robotic engineers, mobile telecommunications, space scientists). Finally, the S3 service will deliver structure detection capabilities addressing the researchers needs to identify and classify, in an efficient way, specific structures of interest.
Within WP4, task T4.1, entitled "Space sector user requirements, service co-design and gap analysis", analyses the data and computational requirements collected by the Space environment for the development of the services. It will further identify existing (and to be developed) EOSC Hub service modules to support the desired/required Space Services functionalities.

1.2. Contents and Rationale

This deliverable, D4.1, is a report containing the requirements, specifications and software development plan for the Space Research Services.
Preliminary documents have been prepared for each of the foreseen Space service (S1, S2, S3), within the NEANIAS WP4 workspace[1], drafting the required datasets and software as well as collecting all possible contributions as Key Performance Indicators (KPIs) to evaluate their success.
The report describes the end-user communities approached to properly collect requirements, the current status of employed software and services as well as the involved datasets and products. Then, it presents the mapping of the requirements with technical specifications, and finally, the software development plan for the implementation of S1, S2 and S3.

1.3. Structure of the document

The deliverable is structured in five sections (chapters), being the first one responsible for introducing the report, briefly mentioning its content and how it is structured. Chapter 2 describes the Space end-user communities (section 2.1), their datasets and products (section 2.2), the current state of the software and services employed (section 2.3) and the

[1] https://citegr.sharepoint.com/:w:/r/teams/h2020-neanias/Shared%20Documents/WP04-Space/02_Services%20Requirements (access limited to NEANIAS members)

requirements (section 2.4), while Chapter 3 details their co-design and specifications, including a gap analysis. Chapter 4 presents the software development plan and guidance, including the presentation of the preliminary architecture of the Space Services and the datasets to validate the services. Finally, Chapter 5 outlines the conclusions and future steps.

2. User Requirements for the Space Services

2.1. Space End-user Communities

Space end-user communities within NEANIAS WP4 (see Table 1) mainly include astrophysics and planetary scientists as well as computer scientists and software engineers interested in computer vision and machine learning. Service data and products may also have a high impact in planetary mining and robotics, space weather and mobile telecommunications, as shown in Table 1.

Space Community	Brief Description
Astrophysics scientists	Study the birth, life and death of stars, planets, galaxies, nebulae and other objects in the Universe.
Planetary scientists	Study planets (including Earth), moons, and planetary systems and the processes that form them.
Planetary mining engineers	Evaluate extra-terrestrial resources available in planets and moons of our Solar System.
Planetary robotics	Operate scientific payloads on remote environments through automation and robotics systems.
Computer vision/ machine learning software engineers	Gain a high-level understanding from digital images and to make automatic predictions through software development.
Mobile telecommunications	Analyse the impact of frequency bands used by Radio Astronomy in standard mobile telecommunications.
Space weather	Study the time-varying conditions within the Solar System, concretely the space surrounding the Earth (including its conditions in the magnetosphere, ionosphere, thermosphere, and exosphere).

Table 1 Space end-user communities

Below, the most represented groups of end-user communities within NEANIAS WP4 are analysed in detail to contextualize the deliverable main themes focused on the space research data, products and software toward the definition of the user requirements to guide the NEANIAS services developments. NEANIAS space services, taking advantage of the unique features of EOSC, will allow tackling the identified challenges, taking astronomy and planetary sciences to a new level, leading to cutting edge discoveries that will change our understanding of the cosmos, and, in short, paving the ground for the science of the future.

2.1.1. Astrophysics Scientists

Astrophysics scientists employ the principles of physics and chemistry to ascertain the nature of astronomical objects, i.e. the Sun, other stars, galaxies, extrasolar planets, the interstellar medium and the cosmic microwave background. They examine the emissions from these objects across all parts of the electromagnetic spectrum, namely their luminosity, density, temperature, and chemical composition. Because astrophysics is a very broad subject, astrophysicists apply concepts and methods from many disciplines of physics, including classical mechanics, electromagnetism, statistical mechanics, thermodynamics, quantum mechanics, relativity, nuclear and particle physics, and atomic and molecular physics. In NEANIAS, a special focus is on Radio astronomy, that studies radiation with a wavelength greater than a few millimetres, and on Infrared astronomy, that studies radiation with a wavelength too long to be visible to the naked eye but shorter than radio waves.

2.1.1.1. Radio Astronomy

Radio astronomy studies focus on sources emitting radiation at wavelengths greater than a few millimetres. These sources could have a Galactic origin, such as clouds of atomic hydrogen (HI regions, see Fig. 1), stars with their surrounding nebulae, usually formed in their final stages (e.g. Luminous Blue Variables, Wolf-Rayet, Planetary Nebulae, Supernova Remnants, etc.), or star-forming regions. At radio wavelengths, one can also observe distant galaxies, in particular, starburst Galaxies and Active Nuclei Galaxies. The observation of such sources is a key tool for understanding the structure of the Universe up to high redshift; but also, for detecting the cosmic background radiation (i.e. the light coming directly from the Big Bang). The detection of such long waves requires telescopes with large diameters or interferometers, which can have their antennas separated by thousands of kilometres.
Very soon, the Square Kilometre Array (SKA[2]), the largest radio interferometer ever built, will start its observational activity to survey the sky, expected to revolutionise our knowledge of the Universe. Indeed, SKA will map the sky with an unprecedented level of detail, reaching ~nJy sensitivity, sub-arcsec spatial resolution and full frequency coverage from 50 MHz to 15 GHz. Moreover, its fast scanning speed will allow the complete mapping of the sky in a short time that was not achievable with previous facilities. To reach such high performances, it will combine the signals received from thousands of antennas spread over South Africa and Australia, simulating a single giant radio telescope. It will be able to survey the sky more than ten thousand times faster than before, requiring the flexibility and scalability of cloud computing resources: it is expected that, at its peak performance, SKA will produce 10 PB per hour.

2.1.1.2. Infrared Astronomy for Star Formation studies in our galaxy

Until the last century, scientists have studied our galaxy, the Milky Way, with optical telescopes in the visible-light, being affected by the presence of clouds of dust in our galaxy. Dust does not allow scientists to observe with these instruments the formation of stars, a process that takes place in the deep and obscured interior of molecular clouds. On the contrary, the infrared radiation can penetrate this screen of dust and allow the direct observation of the surroundings of the new-born stars and still-forming “protostars”. Moreover, the material actively involved in star formation emits mostly in the infrared spectrum, since it heats up to temperatures in the range of 20 to 1500 K. The infrared emission is a probe of the physical conditions of this material, thus becoming the ideal tool to study the process of star formation.

[2] https://www.skatelescope.org/

In the last three decades, there were multiple space missions to observe at infrared wavelengths. From the IRAS[3] mission up to the more recent Spitzer[4] and Herschel[5] space telescopes, astronomers have acquired large amounts of data with progressively improving sensitivity and spatial resolution. These data are showing us in detail the star formation in our galaxy, and have an extraordinary legacy value with a strong potential of serendipitous science that is still being exploited even years after their observations. On the other hand, the infrared window will be still explored with new instruments installed or planned for major ground-based facilities, like the Very Large Telescope (VLT) and the European-Extremely Large Telescope (E-ELT), and space-born, with the forthcoming James Webb Space Telescope (JWST) telescope.

2.1.2. Planetary Scientists

Planetary science is the study of planets and planetary systems composed by moons, ring systems, and magnetospheres. Planetary scientists, then, work on understanding how such systems formed and how they have evolved. It is a cross-discipline field including aspects of astronomy, atmospheric science, chemistry, and geology, for instance.

In NEANIAS, we focus on the geology of solid Solar System bodies – planets and moons. Geologists rely largely on imaging data to recover morphological information over the surface of planetary bodies, while spectral data provide information about the composition of the planets. That information altogether, mixed with evolutionary models, may bring information about the lithology of the bodies. When direct subsurface imaging is not available, it uses non-imaging data as well, like geophysical time-series and potential field data (such as gravity), that indirectly inform on the subsurface of planets and moons (such as in the cases of sounding radar data on the Moon and on Mars).

2.2. Space Datasets and Products

2.2.1. Data and Products for Astrophysics Scientists

In Astrophysics, data are commonly described using the Flexible Image Transport System[6] (FITS). FITS is an open standard defining a digital file format useful for storage, transmission and processing of data: formatted as multi-dimensional arrays, for example 2D images or tables. The FITS standard was designed specifically for astronomical data and includes provisions such as describing photometric and spatial calibration information, together with image origin metadata. FITS was designed with an eye towards long-term archival storage, always ensuring backward compatibility.

3 https://www.jpl.nasa.gov/missions/infrared-astronomical-satellite-iras/
4 http://www.spitzer.caltech.edu/mission
5 https://sci.esa.int/web/herschel
6 https://fits.gsfc.nasa.gov/

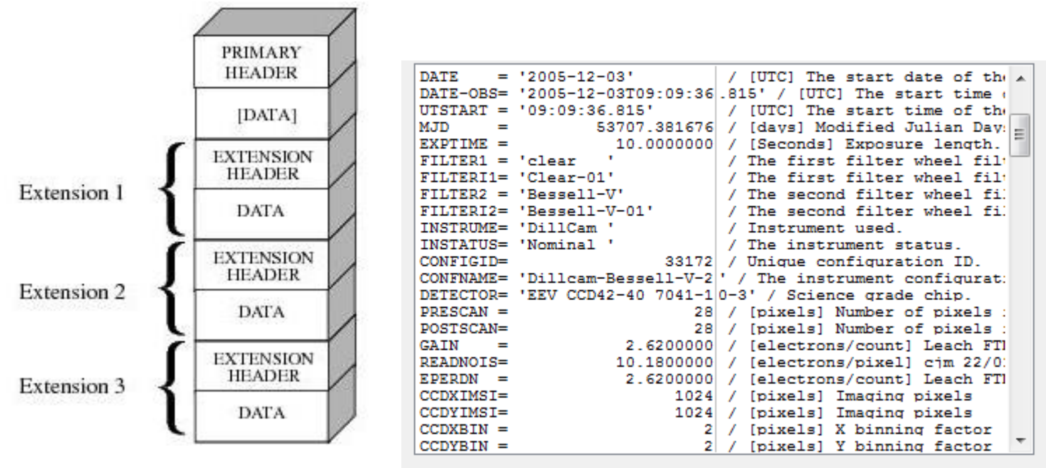

Figure 1 Example of a FITS file structure (Left). Example of the content of a FITS Header, where keywords with info about the image are stored.

Image metadata is stored in a human-readable ASCII header. The information in this header is designed to calculate the byte offset of some information in the subsequent data unit to support direct access to the data cells. Each FITS file consists of one or more headers containing ASCII card images that carry keyword/value pairs, interleaved between data blocks. The keyword/value pairs provide information such as size, origin, coordinates, binary data format, free-form comments, history of the data, and anything else the creator desires: while many keywords are reserved for FITS use, the standard allows arbitrary use of the rest of the name-space.

FITS is also often used to store non-image data, such as spectra, photon lists, data cubes, or structured data such as multi-table databases. A FITS file may contain several extensions, and each of these may contain a data object. For example, it is possible to store x-ray and infrared exposures in the same file.

Astrophysicists mainly deal with astronomical catalogues that are lists or tabulation of astronomical objects, typically grouped together because they share a common type, morphology, origin, means of detection, or method of discovery. The oldest and largest are star catalogues. Hundreds have been published, including general ones and special ones for such items as infrared stars, variable stars, giant stars, multiple star systems, star clusters, and so forth. The most complete in terms of number of objects and reliable distance estimations is currently the one from the Gaia spacecraft (see section 2.2.3.1), that includes measurements for more than a billion stars.

Astronomical catalogues may be compiled from multiple sources, but most modern catalogues are the result of particular astronomical surveys of some kind. Since the late 20th century, catalogues have been increasingly often compiled by computers from an automated survey, and published as text files rather than on paper.

2.2.1.1. Radio Astronomy

Raw interferometric radio data consist of tables corresponding to specific signal correlations for each pair of antennas. The processing of radio data is a complex process that involves several steps:

1) Data flagging: First, the data is flagged to remove faulty records affected by radio frequency interferences (RFI) or instrumental issues (such as receiver failures).
2) Data calibration: Data needs calibration to correct for effects that may hinder its scientific interpretation. Such effects, sometimes frequency-dependent, may be related to instrumental issues (e.g. receiver temperatures) or environmental conditions (e.g. atmospheric fluctuations). Calibration is, therefore, crucial to compare data with other observations and model predictions reliably.
3) Image restoration: Once the data is correctly calibrated, 2D images (for continuum data) or 3D cubes (for spectroscopic data, i.e. with a frequency axis) of the observed patch of the sky are produced by means of dedicated software (e.g. MIRIAD [13], CASA [14]). The generation of these images involves the so-called "deconvolution", i.e. the application of a set of mathematical operations (tightly related to the Fourier Transform) on the data to restore the 'actual' brightness distribution.

The final products of this process are classical FITS files, ready to be analysed and exploited, and thus suitable for many scientific purposes, from the characterization of individual sources (e.g. flux measurements) to catalogue generation.

In NEANIAS, we will use data from public and proprietary surveys, e.g. SGPS [15] or CORNISH [16] and SCORPIO [17], covering different radio bands. The possibility to start working on SCORPIO data is one of the greatest advantages of this project: we will have the opportunity to work on data coming from the ASKAP[7] telescope, the Australian pathfinder of the SKA project. SCORPIO was indeed acquired as part of the early science phase of ASKAP, observing in three bands (from 0.9 to 1.6 GHz) and using the 36 full-set of antennae, thus reaching the optimal sensitivity foreseen for this telescope. Thus, this project will present the same challenges of next generation surveys (i.e. ASKAP EMU[8]; SKA; etc.), both from the technical (data reduction in case of complex, crowded field, i.e. the Galactic Plane) and scientific (e.g. extraction, and classification of compact and extended sources) point of view, becoming the perfect testbench for our software and algorithms. Finally, SKA Science Data Challenges will be considered as well. Those are regularly issued to the community as part of the SKA science preparatory activities. The purpose of these challenges is to inform the development of the data reduction workflows, to allow the science community to get familiar with the standard products the SKA will deliver, and optimise their analysis pipelines to extract science from them. These challenges may consist of real data from currently operating radio facilities or simulated SKA data.

7 https://www.atnf.csiro.au/projects/askap/index.html
8 https://www.atnf.csiro.au/people/Ray.Norris/emu/index.html

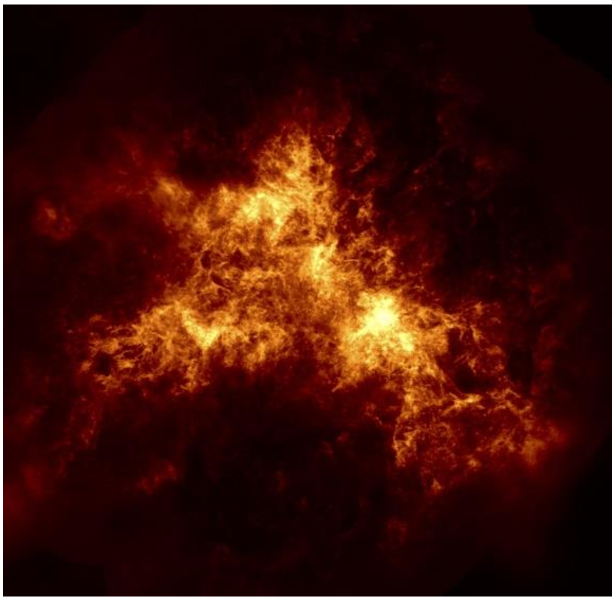

Figure 2 Example of visualisation of a Radio extended source: the atomic hydrogen gas in the Small Magellanic Cloud is imaged with CSIRO's Australian Square Kilometre Array Pathfinder (ASKAP). Credit: ANU and CSIRO

2.2.1.2. Infrared Astronomy data for Star Formation studies in our Galaxy

The infrared datasets used for the scientific studies are mainly of two types: maps and catalogues. The original datasets acquired by the telescope are typically not available to the users and in many cases are even stored in the archive of the survey. These *lev-1* data are calibrated, flagged and projected on the footprint covering the portion of the sky observed, producing the science-ready maps that are considered by the astronomers the raw data for their analysis. Most of the infrared data are 2D maps in the FITS format described above, with the relative metadata saved in the image header, that contains basic information relative to the map, like position of the sky and wavelength, observed, flux calibration adopted, etc. The infrared maps adopted for the studies on star formation are a representation of a portion of our Galaxy as observed by a telescope. Maps of the same region of the sky can appear differently, depending on the diameter of the telescope, the capabilities of its instruments and the observed wavelength.

The catalogue of the sources are typically ASCII tables, in CSV or IPAC standard format, with each row referring to a single astrophysical object, identified and extracted on the infrared maps. The object properties, like position and fluxes at different wavelengths, are listed in the various columns of the table. Several catalogues, but not all, are delivered with a set of metadata that, for example, include practical information on how the catalogues were prepared or collected. The metadata, where available, are particularly useful to reproduce independently the values reported in the catalogues, but they also allow to better match independent catalogues. The most common catalogues list point-like sources, but some are dedicated to extended structures, such as molecular clouds, filaments or bubbles.

The data included in NEANIAS for the scientific studies and demonstration of technical tools and software belong to the main infrared surveys of the Galactic Plane, such as MIPSGAL (covering the Galactic Plane in the longitude range – 60 deg < l < 60 deg)[18], AllWISE [19, 20] and Hi-GAL (covering the entire Galactic Plane) [21]. An example of the image dataset is given in Figure 3. They also include datasets from surveys at different wavelengths, like CORNISH [16] at radio-wavelengths, and spectroscopic surveys of molecular line tracers such as CO isotopologues, the FCRAO Galactic Ring Survey (GRS) [22] and Three-mm Ultimate Mopra Milky Way Survey (ThrUMMS) [23], covering most of the I and IV quadrant of the Galactic Plane. The molecular line data are 3D maps in FITS format that include a further dimension to the images described above. The third axis refers to the velocity shifts with respect to the frequency at rest of the observed species and sometimes can be connected to the distance of the field through the adoption of a Galactic rotation model.

Figure 3 RGB image of a region of the Galactic plane 4 deg x 2 deg wide observed by Herschel Space Observatory in the framework of the project Hi-GAL. The image is obtained combining the maps observed at three different wavelengths, 70 um (blue), 160 um (green) and 250 um (blue) and show two main star-forming complexes and several spots where stars are actively forming (indicated by the blue emission). Image Credit: FP7 ViaLactea Project. Original data from Hi-GAL survey.

2.2.2. Data and Products for Planetary Scientists

Remote-sensing experiments on robotic automated spacecraft are the preferred source of data used for most of planetary geoscience. Thus, image arrays, or more precisely, *rasters* are the predominant data format. Broadly used are GeoTIFF for the storage of multi-layered georeferenced bi-dimensional arrays (aka, *raster* data), and vector data such as ESRI Shapefiles for storing georeferenced polygonal, linear or point elements along their properties, generically called *vector* data. Fairly recently Geopackage[9] was developed as an open standard for storing vector data (mainly) in a more concise way compared to Shapefiles or other vector formats.
When it comes to modern planetary missions the set of storage formats is incremented to support a greater complexity of data and the processing tools. Data from the CRISM

9 https://www.geopackage.org/

instrument, onboard of the Mars Reconnaissance Orbiter [25], providing high-resolution pixel-wise spectra, are distributed in multidimensional cubes, which can use several file formats, such as ENVI .img (+hdr), FITS, or ISIS3 cube (.cub) formats, storing multi-band, multi-dimensional raster datasets.
Although NEANIAS systems must consume different data formats – according to the origin of data being processed – the output provided to our users will span a narrow set of formats so that we focus on providing a homogeneous interface to optimise data accessibility and to satisfy open standards supported by the Open Geospatial Consortium (OGC), e.g., with GeoTIFF, or the International Virtual Observatory Alliance (IVOA), e.g., with VOTable or FITS files, towards *accessibility* and *interoperability* after the FAIR (Findable, Accessible, Interoperable, Reusable) guidelines [27], a major guideline of the NEANIAS project.

The following protocols and data formats will interface and deliver data to users:

- GeoTIFF (OGC)
- GeoPackage (OGC)
- WMS, WFS, WCS (OGC)
- EPN-TAP (IVOA)

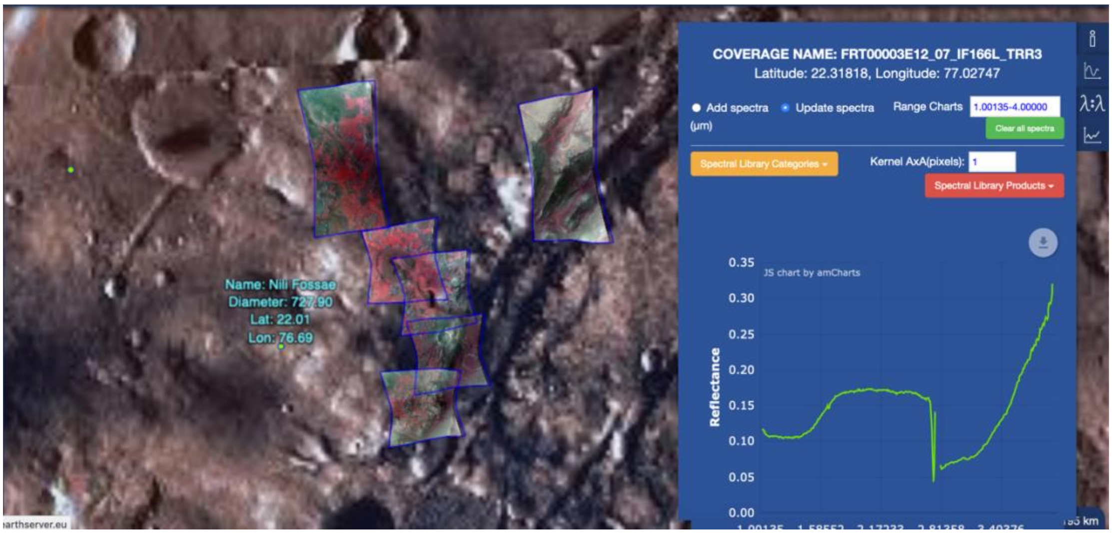

Figure 4 Exemplary raster imagery over planetary surface, RGB color combinations extracted from hyperspectral cubes (source PlanetServer)

2.2.3. Other Data and Products

Apart from the aforementioned data, the GAIA catalogue will be employed in NEANIAS WP4 for Virtual Reality navigation to benefit the identified space user-communities. The GAIA spacecraft[10] is measuring the positions, distances and motions of stars with unprecedented precision, constructing the largest and most precise 3D space catalog ever made, including

10 https://sci.esa.int/web/gaia

billions of astronomical objects, mainly stars, but also planets, comets, asteroids and quasars among others.

GAIA catalogue data release 2 and 3 are the last releases of catalogue available at DPCT, in ALTEC (WP4 partner). The mission was extended until December 2020, if new extension will be decided other two catalogue releases shall be stored at DPCT. The DPCT shall keep online the last two versions of the catalogue so versions 2 and 3 are available in the Relational Data Base Management System (RDBMS) and files. Older catalogue versions are available as files. The last catalogue release contains about 8 billions of sources reporting astrometric and spatial information. Having on online catalogue in an RDBMS allows to research data exploiting spatial criteria too. The spatial research is made available creating 'ad hoc' structure to maintain not altered the original data. However, enabling search on RDBMS have a cost in term of space, the GAIA catalogue size is about 24TB respect to compressed file size that is 1.5TB.

2.3. Space current Software and Services

This section reports on the state of the art of the current software and tools (in TRL6) in use within the Space Research community.

2.3.1. Astrophysicists Software and Services

2.3.1.1. VisIVO and VLVA

The ViaLactea Visual Analytics (VLVA) tool [1], based on the VisIVO suite, is aimed at exploiting a combination of new-generation surveys of the Galactic Plane to study the star formation process of the Milky Way. It combines different types of visualisation to perform the analysis exploring the correlation between different data, for example 2D intensity images with 3D molecular spectral cubes. The scientist is equipped with a tool for discovering the link between different physical structures, from the extended filamentary-shaped structures to the most compact, dense sources precursors of new stars. The implementation philosophy behind VLVA is to provide a means to make transparent to scientists the access to all underlying information without requiring technical skills, e.g. to query the VLKB (ViaLactea Knowledge Base). Such ways of providing easy accessibility to huge amounts of data is offered from the very first user interaction with the tool that helps to identify candidate star formation sites within a panoramic view of the Milky way.
In NEANIAS the VLVA tool will be extended to exploit the new cloud opportunities given by the VLKB for FAIR data management and enlarging analysis techniques on the cloud.

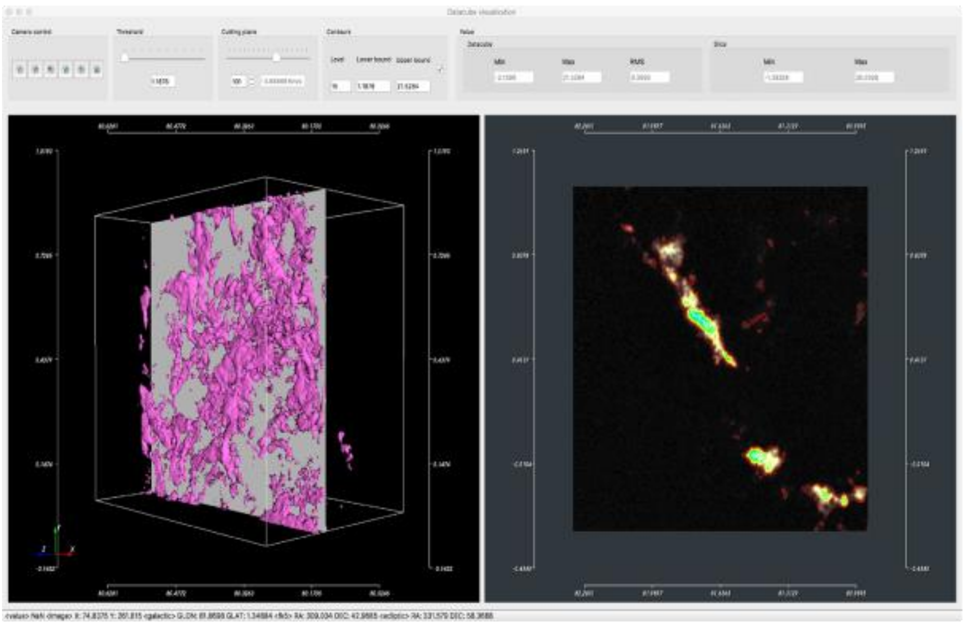

Figure 5 Sample screenshot of the ViaLactea Visual Analytic tool.

2.3.1.2. VLKB

The ViaLactea Knowledge Base (VLKB) [2] includes a combination of storage facilities, relational databases and web services on top of them. It collects heterogeneous data such as observational data, in the form of images or multi-dimensional datasets and catalogues, alongside database schemata and relations dedicated both to meta description of the mentioned data collections and catalogues. The discovery, access and retrieval solutions on top of these resources are based on an IVOA TAP service for all of the database content that needs to be exposed to the community plus some dedicated search, cut-out and merge solutions for the 2D and 3D datasets available through the project. This latter secures the datasets (which are a mix of public and private policy ones) and allows searching also within novel datasets, such as the proprietary ViaLactea datasets on 3D extinction maps of the galaxy. The underlying TRL6 methodology and algorithms will be deployed for the open science datasets such as the ones to be used within NEANIAS.

2.3.1.3. VR and ADN

Astra Data Navigator (ADN) is a tool for 3D visualisation of stellar catalogues (stars, asteroids, planets, satellites, etc ...) in a navigable environment.
The creation of the models takes place in real time based on the physical parameters read from the catalogues to try to make the representation as realistic as possible.
In addition to reading the stellar catalogues, for bodies where it is possible, the placement in the environment is carried out through the use of Spice Kernels.
ADN software have been tested on the Hipparcos datasets and will be extended to use Gaia datasets.
Currently ADN loads catalogues from files (one as main catalogue to generate the star, Hipparcos contains only stars, and other files to load planets, satellites, etc...).
This capability is provided through a specific implementation of a common interface.
Is therefore possible to create implementations of same interface for reading catalogues from other sources such as databases. This approach is required to allow for system scalability with the least possible impact on performance. Hipparcos is a catalogue with about 110K entries while GAIA has 2B of entries therefore reading from files, event with a linear complexity, may require suboptimal loading times.

2.3.1.4. MONTAGE

Montage is a command-line toolkit containing modules for reading a collection of FITS images and performing all the processing steps needed to assemble them into a mosaic. It has been used to generate mosaics from data released both by space-based telescopes, e.g. Spitzer Space Telescope, Hubble Space Telescope, the Infrared Astronomical Satellite (IRAS), the Midcourse Space Experiment (MSX), the Sloan Digital Sky Survey (SDSS), and ground-based telescopes such as the National Optical Astronomy Observatories (NOAO) 4-m telescope and the William Herschel 4-m telescope, thus using data from a wide wavelength range.
Its scientific value derives from three features of its design: 1) it uses algorithms that preserve the calibration and positional (astrometric) fidelity of the input images to deliver mosaics that meet user-specified parameters of projection, coordinates, and spatial scale; 2) it contains independent modules for analysing the geometry of images on the sky, and for creating and managing mosaics; these modules are powerful tools in their own right and have applicability

outside mosaic production, in areas such as data validation; 3) it is written in American National Standards Institute (ANSI)-compliant C, and is portable and scalable - the same engine runs on desktops, clusters or supercomputer environments running common Unix-based operating systems.

2.3.1.5. UNIMAP

UNIMAP [3] is a map maker designed to work on the data from PACS and SPIRE, the two infrared photometric instruments installed on the Herschel Space Observatory, for observations in scan-mode of regions of the sky. It processes directly the raw *lev-1* data, i.e. the detector readouts converted and calibrated into physical units, to produce high-quality maps. UNIMAP is based on a Generalized Least Squares (GLS) technique to correct the temporally correlated noise typical of the detector of these instruments. However, it is designed to also address at the same time the two major issues present on Herschel datasets that other map-maker algorithms were not able to treat properly: the spatial correlation of the noise due to adjacent elements of the detector and the distortion induced by strong variations of the signal. UNIMAP deals with these issues using two algorithms, an Alternative Least Squares (ALS) and a post-processing GLS (PGLS) algorithm, both implemented and integrated in its package. The software package is written in Matlab, but it does not require Matlab to run. Compiled versions are available for Mac and Linux.

2.3.1.6. CAESAR

CAESAR [4,5] extracts and parametrizes both compact and extended sources from astronomical radio interferometric maps. The processing pipeline is a series of stages that can run on multiple cores and processors. After local background and rms map computation, compact sources are extracted with flood-fill and blob finder algorithms, processed (selection + deblending), and fitted using a 2D gaussian mixture model. Extended source search is based on a pre-filtering stage, allowing image denoising, compact source removal and enhancement of diffuse emission, followed by a final segmentation. Different algorithms are available for image filtering and segmentation. The outputs delivered to the user include source fitted and shape parameters, regions and contours. Written in C++, CAESAR is designed to handle the large-scale surveys planned with the Square Kilometre Array (SKA) and its precursors.

2.3.1.7. CUTEX and ViaLactea Filament Finder

CuTEx [6] is a photometric package designed to process FITS images to identify and extract compact sources embedded in complex backgrounds, like the ones present in star-forming regions or the Galactic Plane observed at infrared/sub-mm wavelengths. Physical objects in these images, such as protostars in their accreting phase, are generally not point-like, so their shape is not fitted by a PSF-profile. CuTEx can detect and extract these objects even in cases of high crowding and in location where the background is highly spatially variable. The package is composed of two main algorithms, one tuned for the source detection, and the other for source extraction dedicate to the measure of fluxes and sizes. CuTEx is written in IDL language and it runs on any machine with an IDL compiler.

ViaLactea Filament Finder [7] (VLFF) is a different package aiming to deliver an automatic detection tool able to identify complex filamentary structures present of infrared 2D FITS maps and to characterize their morphological and physical properties. It is tailored to identify patterns with a precise morphological definition, i.e. elongated, cylinder-like regions with a relatively higher brightness contrast with respect to their surroundings, that describe filaments in the most general and abstracted way. The tool implements different methods to trace these features and determine the relative properties [8], like for example intensity, length, width, spine position and orientation, etc. The package is written in IDL and it is composed by multiple routines called from a main wrapper.

2.3.2. Planetary Science Software and Services

2.3.2.1. PlanetServer (raster-serving, data access and query)

PlanetServer is a web service providing simplified access to hyper-spectral data cubes [12]. The service is composed by a client – graphical, running on a web-browser – and a server, responsible for handling the client requests and data management.

At the interface, the server implements two communications channels to clients, outside world: (i) interactive requests by the graphical interface from the users' web-browser, and (ii) REST API interface answering to low-level requests for data products.

PlanetServer server-side system is composed by several specialist sub-components dedicated to different stages of the data management. Mainly,

- a database, where data is effectively stored;
- a REST API, interfacing stateless public requests;
- a processing pipeline, serving user requests from the query interface to the database and back to the user.

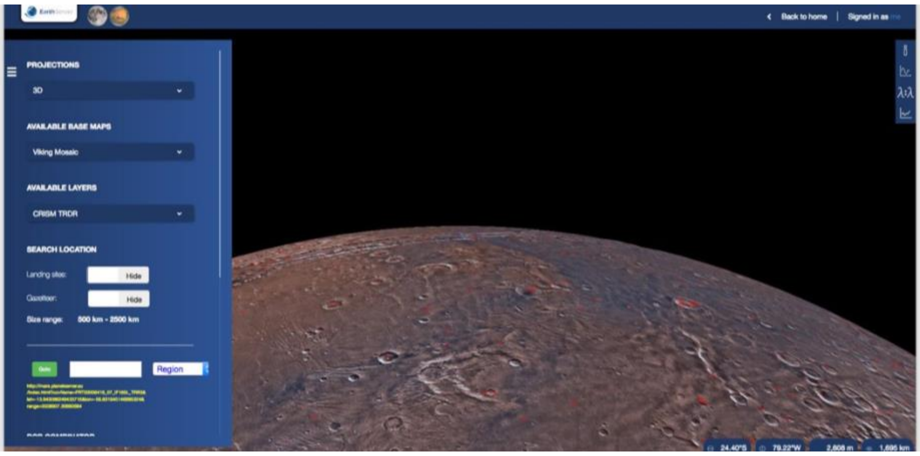

Figure 6 Sample screenshot of the PlanetServer tool.

2.3.2.2. ADAM (raster-serving, data access and query)

The Advanced geospatial Data Management platform (ADAM) is a tool to access a large variety and volume of global environmental data. ADAM allows you extracting global as well as local data, from the past, current time, as well as short term forecast and long-term projections. Most of the data are updated daily to allow users having always the most recent data to play with.

The core of ADAM is a Data Access System (DAS), a software module that manages a large variety of geospatial information that feature different data format, geographic / geometric and time resolution. It allows accessing, visualizing, sub-setting, combining, processing, downloading all data sources simultaneously. The DAS exposes OGC Open Search and Web Coverage Service (WCS 2.x) interfaces that allow discovering available collections and subset them in any dimension with a single query.

ADAM is a modular platform: various DAS are deployed on different data sources (DIAS Mundi, DIAS creodias, Amazon Web Services - AWS, MEEO Data Facility, SISTEMA Data facility), allowing accessing and sub-setting the available collections without downloading / duplicating the data. Distributed data sources are made accessible through the datacube layer, that exposes OGC-standardised interfaces.

ADAM offers three main interfaces for data access:

- the **Explorer**, a web-based graphic user interface to allow users to explore, access, process and download data
- the **ADAM API**, that provide a python-library to directly access the ADAM data access and processing capabilities directly integrated in the user's code and applications
- the **Jupyter Hub**, a web-based processing environment to allow users to import, write and execute code that runs close to the data, exploiting the power and the APIs on a remote computation environment (no user resources are used).

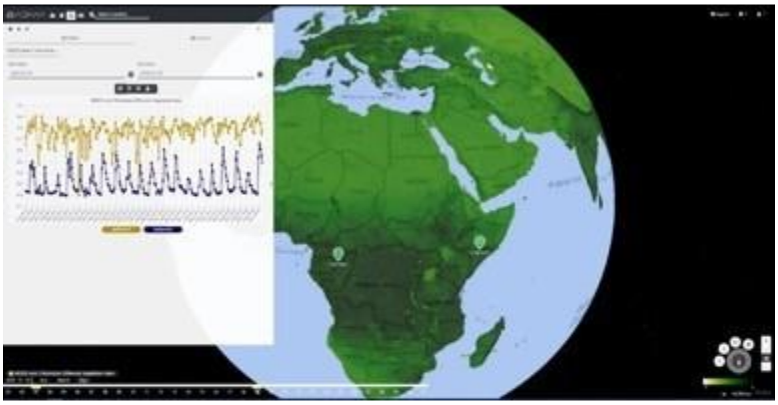

Figure 7 Sample screenshot of the ADAM Explorer.

Besides data exploitation and analytics user interfaces (Explorer, ADAM API, Jupyter Hub), ADAM allows at integrating data processing pipelines in the Data Processing System (DPS). Among all management functionalities, DPS aims at performing:

- Pipelines configuration: Besides the pipeline parameters configuration, the dashboard provides also the possibility to compute the expected cost of the job, based on configurable cost plans. Pipelines based on the open source SNAP libraries are already available at the current stage.
- Jobs management: From each pipeline, many jobs can be launched changing e.g. AOI, start/end date, source data, and so on. For each job, progress status is provided (progress, processed products, instances) as well as estimated termination date/time. A deadline can be set to safeguard the system for overloading. Elastic job parallelization options are also available on cloud providers that allow this possibility via command line interface (e.g. AWS): the horizontal scalability is based on Docker[11] and Kubernetes[12] technologies.

2.3.2.3. USGS-ISIS (processing and mosaicking)

ISIS3 [9, 11] is a set of tools for planetary data analysis. It provides either command-line based tools as well as graphical viewers; some are specific to certain space missions (*e.g.*, for the MRO/CRISM mission) while other tools are meant for general data handling. All together ISIS3 is one of the cornerstones of current geo-planetary data analysis.

ISIS3 is a modular package composed by a *core* module and *mission-specific* modules. *Core* provides the base set of tools, settings and support data for general data handling. On top of that, users can attach specific tools and support data corresponding to one or more missions of interest.

ISIS3 is used internally for the mosaicking pipeline[13]. In NEANIAS we are going to expose the mosaicking functionality for the specific missions of the project. ISIS3 tools portability and scalability should improve to suite a cloud computing environment, according to our expertise, to which Docker containers will be used for encapsulating ISIS3 modules.

As a general guideline, as implemented in PlanetServer the user interface for processing and data retrieval will be composed by two levels of interaction: (*i*) a REST API providing low-level, though well-defined query interface, and (*ii*) a graphical web interface for the user willing to explore the data at disposal.

2.3.2.4. NASA-ASP (stereogrammetry)

ASP [10] is a suite for producing cartographic products such a Digital Elevation Models (DEMs) and 3D models from stereo imagery. Stereo images are images taken from coincident regions but different lines-of-sight, which, with the right tools, allow us to recover information on the third-spatial axis (*I.e.*, elevation). Effectively, with ASP tools we can reconstruct 3-dimensional maps from satellite, as well as robotic rover images from Mars, for instance.

Digital Elevation Models are particularly important for geoscience since elevation information have an impact in other terrain analysis (e.g., lithology, sub-surface structures). In NEANIAS,

[11] https://docs.docker.com
[12] https://it.wikipedia.org/wiki/Kubernetes
[13] https://astrocloud.wr.usgs.gov/index.php

like ISIS3, ASP is going to be ported to the cloud environment to provide automated DEMs production for the selected datasets.

2.3.2.5. Map serving services (products serving)

Processed planetary datasets produced by the users must be properly allocated for open data access. Open data access and FAIR guidelines are to be accomplished by (i) following international standards, (ii) simplifying the interface with the user, and (iii) making every product a public asset, including user requested products.
While for original mission data archives Rasdaman/PlanetServer will be used as the database manager for image arrays and data-cubes, vector data products will be handled by GeoServer[14]as it can store not only raster data array but vector datasets as well, which will be used to serve observation footprints -- from source datasets – as well as surface features outline.

2.4. User Requirements for Space Services

In the first instance, an initial set of user requirements has been formed by using the paradigm of 'User Stories' as recommended by the NEANIAS Agile software development approach in order to capture the description of software features from end-user perspectives. A User Story describes in a short way the type of user, what they need to achieve and why. A User Story is typically very high-level and helps to create a simplified description of a requirement that can be subsequently transcribed into development specifications. Usually, a User Story provides in one sentence enough information related to the described product feature, for which the development team can conduct a reasonable work load estimation. Furthermore, the User Story is used in planning meetings to enable the development team to design and implement the product features.

A User Story typically has a predefined structure like the following:

As a <user-type (stakeholder)>, I want to <user-requirement> so that <reason>

The aforementioned user stories/requirements were linked with a level of priority and value based on end-user recommendations as explained in Section 2.5. Also, following extensive discussions within the user base the final list of prioritized requirements was delivered. The list is not meant to be exhaustive, rather a sound starting point that can evolve over the project.

2.4.1. User Requirements for Astrophysics services

The following table presents the User Stories collected by the Astrophysics community, which are described and motivated in detail in the paragraphs below.

Id	End-User	User-requirement	Reason
RA-1	Astrophysicist	To analyse maps at different wavelengths	To better analyse different emitting components in the

[14] http://geoserver.org

			same astrophysical source from the same UI or service
RA-2	Astrophysicist	To reduce images to the same technical features	To compare images from different surveys
RA-3	Astrophysicist	To enhance current source finding algorithms and software	To identify, classify and characterize compact and extended sources with a minor impact of artefacts and spurious detection.
RA-4	Software engineer	To improve the software's portability	To easily share software and code within the community not being limited to available local computing resources
RA-5	Astrophysicist	To improve the algorithms' reproducibility	To share the acquired knowledge and allow upgraded experiments
RA-6	Astrophysicist	To access all observations and catalogues available for a certain sky region and epoch	To have a dataset to perform data analysis or train and validate analysis algorithms
RA-7	Software engineer	To integrate existing data access and processing components	To create a more complex integrated system capable of producing new scientific information

Table 2 Summary of the Astrophysics User Requirements

Considering the above User Requirements, the following Recommendations are provided.

Recommendation Astro 1: Increase multiwavelength studies.
The new approach in astrophysical studies is the analysis of the same source at different wavelengths, spanning a range from high to low frequencies as wide as possible. In this way, it is possible to better disentangle and analyse different components of the same source, possibly characterized by different physical conditions.
To do that, we need to: 1- access different public surveys and their own archives, being subjected to their specific query formats; 2- manage FITS files with different kinds of technical features (e.g. dimensions, resolution, etc.). A service that offers direct access to all of these data together with the possibility of combining/mosaicking data is absolutely necessary to

speed up the research and offer the scientific community the most complete knowledge database.

Recommendation Astro 2: Reduce images to the same technical features.
As anticipated in Recommendation Astro 1, dealing with different survey images poses a major problem: how to compare images and maps with different dimensions, spatial resolution, pixel size, units for the flux info storage, etc. A tool that manipulates the images in order to make them directly comparable represents a service of key importance for the Community.

Recommendation Astro 3: Improve source identification, classification and characterization.
Many current source finding algorithms and software may detect many false positive sources when processing maps with complex background to generate a catalogue. Ideally, these artefacts and spurious sources should be removed from the final source catalogues. Residual false detections are currently manually removed by visual inspection in almost all surveys. This is error-prone, not reproducible, time-consuming and unfeasible at the scale of SKA. Moreover, existing source finders do not provide the astronomical identity of extracted sources and, finally in the low S/N regime, compact source detection performances are poor, e.g. parameter (flux density and position) estimation are biased due to inaccurate background estimation, deblending and fitting. Machine Learning may help in: i) identifying real and good sources from false/bad sources in radio catalogues automatically produced by existing source finders; ii) identifying astronomical object classes of extracted compact sources; and iii) outperforming detection and characterization capabilities of existing finders, thus improving completeness and reliability.

Recommendation Astro 4: Improve portability, distribution and performance.
Many software packages, code and scripts employed to study astrophysical phenomena are currently executed in local laptops and PCs and shared among few researchers often within the same research group. At the scale of next generation telescopes and data size, it will be impossible to process data using local resources, but computing should approach data to minimise its transfer and replication.

Recommendation Astro 5: Improve reproducibility.
In every science field, the reproducibility of the results obtained is fundamental to prove their own reliability.
Astronomy is producing data at an unmatched rate. The installation of new telescopes, combined with marked improvements in pattern-finding algorithms, has led astronomers to turn to sophisticated software to do the data-crunching they cannot do manually. With more powerful analyses, there is less transparency as to how they have been performed.
The idea is to use open-source code libraries and other Internet resources to publish explanations of what has been done to the datasets, since the moment they were collected, and to make all of them available for peer review.
Once the scientific community can freely glean not only to the same data archive (Recommendation Astro 1), but also to all published software and algorithms, then we will

reach the optimal reproducibility, giving everybody the tools to test the experiment and validate the results, in a sort of world-wide laboratory.

Recommendation Astro 6: Improve data access.
Datasets are described using different metadata, saved with different formats and stored in different locations. In this scenario datasets exploitation is complex and tedious. There is the need to facilitate the data access through a common approach based on a common dataset's description, standards for data format and efficient distributed data stores. The idea that will be carried forward in NEANIAS is to define and design a data access layer that will implement this approach and, while at the same time reusing existing applications and tools integrating them in this data access layer.

Recommendation Astro 7: Improve integrability.
The available applications and tools can be used to provide standalone space services but they could be also combined to create new services that are the result of data access and processing integration. This user story aims at allowing the end-user to compose complex scientific workflows using single space services as element of higher workflow. In order to reach the integrability goal the space service architecture should be designed and implemented to satisfy this technical requirement. REST APIs have to be considered to expose space services in a way that is easily invocable from external systems.

2.4.2. User Requirements for Planetary services

The following table presents the User Stories collected about the Planetary community.

Id	End-User	User-requirement	Reason
RP-1	Geologist	To improve image mosaicking automation.	Mosaicking demands specialist knowledge about data and tools; should be optimised/fully automated to allow real-time data analysis.
RP-2	Robotics engineer	To improve detection of landing sites and landscape routes.	When sending robots to other planets a proper (smooth, flat) landing site and route are required.
RP-3	Mining engineer	To improve lithology prediction.	A precise idea of mineral resources reservoirs is crucial for space/planets exploration planning.
RP-4	Software engineer	To parallelise Rasdaman, array database manager.	Rasdaman community is serial. Bringing Rasdaman to cloud demands parallelisation.
RP-5	Geologist	To improve reproducibility	Open data and open software are crucial for results reproducibility, although it must be improved with auto-generated documentation.

Table 3 Planetary Science User Requirements

In the following paragraphs each User Story has been further described and motivated.

Recommendation Planetary 1: Automate image registration and mosaicking
Combining overlapping images – or partially overlapping – serves to broaden the field of view or enhancement of the signal; or both. Typically, high-resolution images have limited coverage since they are computationally expensive (i.e., time- and storage-consuming), whereas low resolution images provide wide-field coverages. To have the best of both worlds – wide-field

coverage with high-resolution signal wherever possible – mosaicking must be applied. Figure 8 illustrates the what a mosaic data product looks like.
Although it may seem a simple process if we think about matching features in overlapping images, the actual details to be considered and merged, at pixel level, is a multi-dimensional problem to solve: images may have different luminance levels, pixel resolution, noise levels, as well as small distortions because of different lines-of-sight. Consequently, the mosaicking pipeline is a non-trivial algorithm with many parameters.
An automated process may be reached to provide science-ready mosaic images considering a control source dataset.

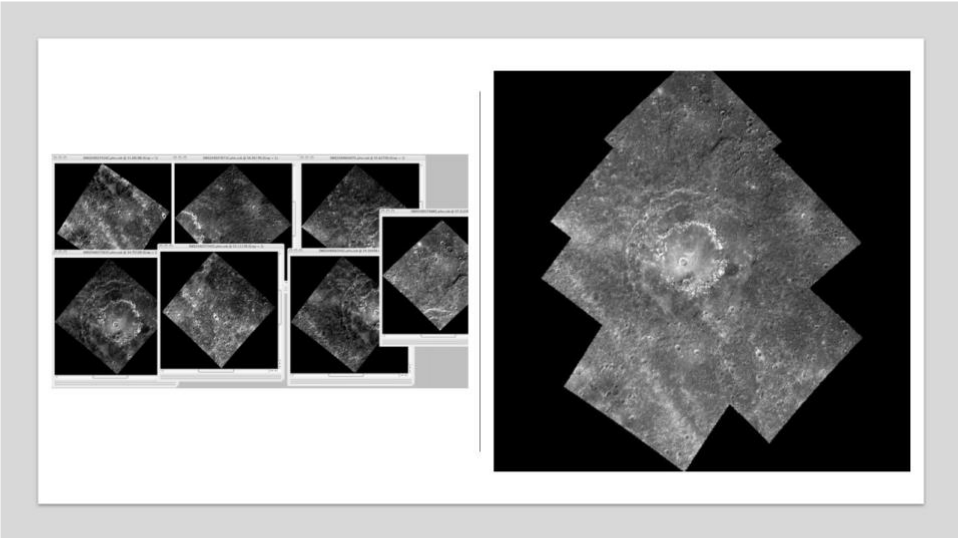

Figure 8 Example of mosaic: input (left) and product (right) [26].

Recommendation Planetary 2: Evaluate and characterise landing sites
Planetary exploration just entered a new era, after mastering satellite missions around planets, moons and even asteroids, we are now greatly succeeding *in situ* exploration with rovers, land-stations and other robots to come. Roughly speaking, such missions are composed by two parts, where the first concerns the launching and the journey until the target planet, while the second is about landing and performing the route through the mission sites.
The landing sites are defined according to the mission's goals; a rover robot, for instance, has to land in a place that (1) is flat, and (2) from where it can follow a route through a smooth path, considering that the meaning of "smooth" is defined according to the rover's capabilities.
It would be valuable to a broad audience of engineers, experts in robotics but not necessarily in planetary geosciences, to have a semi-automated tool to evaluate the smoothness – or, opposite, the roughness – of an extra-terrestrial body given its satellite imagery.

Recommendation Planetary 3: Evaluate, map and predict mineral resources
Space exploration has been getting substantial attention from the private sector during the last decade. Space flight engineering has gotten mature enough to make space flights safe, and now with the help of private sector, they have become cheaper; with cargo flights to the

International Space Station nowadays a daily routine. Such progress inevitably pushes our thoughts towards resources, what used to be science-fiction are today's plans and will soon become a reality.
The detection, or more precisely, the prediction of mineral resources reservoirs is a key component in this human exploration plan as the identification of such resources -- and their respective volumes estimates – should influence the scheduling of future mission deployments.

Recommendation Planetary 4: Parallelise Rasdaman database manager
Rasdaman[15] is one of the pillars of PlanetServer, besides being a promising database software in the array-databases landscape[16]. Not only it provides the provides storage and management/query interface for data format natural for geosciences, but Rasdaman also provides processing capabilities in real-time through its WCPS[17] interface. Rasdaman needs nevertheless to be deployed on to the EOSC cloud, which demands development towards scalability in a distributed system.

Recommendation Planetary 5: improve reproducibility
The Scientific method is based on a handful of guidelines, one of them is *reproducibility*: (virtually) everybody should be able to reproduce a claimed scientific result, for it can be considered a scientific result.
Recently, thanks to the massive incorporation of software into the scientific workflow and the big volumes of data, the scientific community returned to the discussion about reproducibility of results; despite data and software privacy now being considered as a potential blocker. Open data and open software are basic aspects of the modern scientific method; they are crucial not only for science reproducibility but also for science sustainability. To reach reproducibility in a software-based laboratory it is necessary to instrument the whole processing chain, from data source and input parameters all the way down to results, where "instrumenting" means to track the data processing stages and the metadata defining the environment around.

2.5. User Requirements Level of Priority, Value and Acceptance

Finally, we present a table linking user stories and requirements with a level of priority and value based on User Board feedback[18] (see Appendix for additional details) and end-user recommendations. Also, the manner to confirm their acceptance is detailed.

15 https://www.rasdaman.com/
16 https://rd-alliance.org/system/files/Array-Databases_final-report.pdf
17 https://www.opengeospatial.org/standards/wcps
18 https://citegr.sharepoint.com/:f:/r/teams/h2020-neanias/Shared%20Documents/WP04-Space/01_WP04%20Deliverables/01_D4.1_M6/UserBoard_FeedBack?csf=1&web=1&e=o7lgFh
(access limited to NEANIAS members only)

Id	Priority	Value	Acceptance
RA-1	High	8	Availability of different surveys
RA-2	Medium	6	The tool is available to process the multiwavelength data
RA-3	High	9	Use of automatic source finding tools
RA-4	High	10	Use of portable technologies such as VM and Containers
RA-5	Medium	8	Data, software and workflow available to reproduce results
RA-6	High	10	Homogeneous data access layer
RA-7	High	9	Available framework to integrate different services (e.g. REST-API)
RP-1	High	8	Able to produce (mosaic/field) images by selecting the source images and optional parameters (i.e., automated process).
RP-2	Medium	9	Able to filter regions in source images based on geographic and morphometric parameters: lat, long, topographic height.
RP-3	Medium	8	Able to detect minerals on custom/new data cubes by using pre-defined templates.
RP-4	High	7	Balanced data ingestion and queries throughout a cluster of machines.
RP-5	High	9	Data, software and workflow available to reproduce results

Table 4 User Requirements Level of Priority, Value and Acceptance

3. Co-design and Service Specifications

3.1. NEANIAS Space Research Services

This section presents the NEANIAS Space Research Services as identified from the WP4 toward defining the technical specifications, architecture and software development plan.

3.1.1. S1 - FAIR Data Management and Visualisation for Complex Data and Metadata

The Space environment shares different needs ranging from management of complex datasets, scientific visualisation, visual analytics including combining data analytics and mining with visualisation also exposing outputs within advanced interaction environments using virtual and augmented reality. The developed tools and services in Space are tailored for 2D and 3D model computation, visualisation and virtual reality navigation applications. VisIVO and VLVA are suites of visualisation software and services targeting the protocols and standards of the International Virtual Observatory Alliance (IVOA) and exploiting High Performance Computing and distributed computing infrastructures. The ADN virtual reality application allows realistic and precise visualisation of astronomical catalogues in a virtual, navigable and interactive 3D environment. PlanetServer allows planetary surface data access via client/server interactions and 2D/3D visualisation. These tools and services are all in TRL 6 status and will be ported to the EOSC, starting from experiences in previous projects (e.g. in the EOSCPilot project), by optimising specific features as detailed by the involved scientific groups in Section 2.4.

The S1 service will optimise data discovery of VLKB under the FAIR principles by the publication of outputs through the IVOA framework, which is in very wide use by the astronomical community and is increasingly being adopted by the planetary community following the EuroPlanet, VESPA and PlanMap initiatives. Data accessibility will be guaranteed through the deployment of standard protocols and interfaces implemented by the IVOA TAP and EPN-TAP protocols and the REST interface.

3.1.2. S2 - Map making and mosaicking for multidimensional images

Space science requires services for making high quality images from the raw data captured by instruments (map making) and for assembling those images into custom mosaics (mosaicking). The tools available for the community such as Unimap and Montage, ISIS3 and ASP (that are in TRL 6) will be ported to the EOSC, optimising the specific features as needed by the relevant scientific groups. Concerning radio-astronomy data, interferometric observations, due to the peculiarities of data acquisition, the NEANIAS S2 service will consider primary beam effects as a weight for data to be mosaicked. Planetary mapping requires high image quality so that individual data products/granule are co-registered to achieve iteratively cartographic description (e.g. geologic) of a planetary surface. To achieve such mapping, there is the need for higher-level products (high-level data are not always available on public space agency archives, such as PDS/PSA) made available, co-registered and mosaicked as required. An ad-hoc instance of ADAM Space will be deployed to support the Planetary Science user community. The ADAM components and functionalities will be extended: i) to enable data-cube access services on top of planetary science data; ii) the design and implementation of specific data drivers (e.g. CRISM) and data preparation pipelines; iii) to customize the Explorer frontend to discover, access, explore the Planetary Science datasets; iv) to integrate the EOSC authentication service.

3.1.3. S3 - Structure detection on large map images with machine learning techniques

The advent of new-generation telescopes and all-sky surveys has caused a dramatic increase of Space data volumes, making automatic structure detection a necessity. This need will become more and more urgent due to the continuous increase of data volumes to be analysed. Recently, project partners were involved in projects for automatic structure recognition such as FP7 ViaLactea project and SKA precursor activities. CAESAR is a software tool for extraction and parameterization of both compact and extended sources present in astronomical maps. CuTEx and VLFF analyses images in the infrared bands and, in particular, they were designed to resolve problems concerning the study of star forming regions. Planetary exploration missions have the constant need for terrain characterisation, largely based on orbital remote sensing data. S3 service will be focused on exploiting the existing TRL6 software to perform pattern and structure detection in astronomical surveys as well as in planetary surface composition, topography and morphometry.
The S3 service is expected to integrate in EOSC cutting-edge machine learning (ML) and deep learning (DL) algorithms, developing convolutional neural-networks for computer vision tasks (e.g. recognition, segmentation) adapted to the project-specific tasks by means of transfer learning approaches, to perform advanced classification for structures of sources in the sky or planetary surfaces to identify regions of interest.

3.2. NEANIAS Space Service Co-design Specifications

This section summarises the Space community user-requirements linking to the specifications towards porting the NEANIAS Space services within EOSC.

	Requirement				Specification			
Community	**Req#**	**Requirement**	**Requirement Type[19]**	**Rank[20]**	**Current solution**	**Gaps**	**Proposed improvement in NEANIAS**	**Related Space Service**
Astrophysics	**RA1**	Increase multi wavelength studies	Storage	Mandatory	Using VLKB and VLVA	Currently missing some surveys	Increase storage capabilities on the EOSC under FAIR principles	S1, S2
	RA2	Reduce images to the same technical features	Computing	Optional	Using Montage	Currently done on local computing resources	Increase computing capabilities in EOSC	S2
	RA3	Improve source finding	Computing	Mandatory	Using CAESAR , CUTEX and VLFF	Currently missing ML/DL capabilities	Use of powerful computing resources for DL/ML (e.g. GPU)	S3
	RA4	Improve portability	Cloud	Mandatory	Using Singularity container for CAESAR tool	Currently only few software/services available on VM/container technologies	Increase use of virtualization and container technologies for easy portability	S1,S2,S3
	RA5	Improve reproducibility	Cloud	Convenient	Using VLVA and CAESAR, CUTEX and VLFF	Currently sharing bash scripts	Increase use of Jupyter Notebooks for easier reproducibility and accessibility to the software	S3
	RA6	Improve Data Access	Storage/ Cloud	Mandatory	Using VLKB and custom data stores	Standalone applications and only partial availability of web services	Improve dataset description, standard usage, storage on cloud and data access API	S1
	RA7	Improve integrability	Computing/ Cloud	Convenient	Using CAESAR , CUTEX and VLFF, Montage	Currently standalone processing and analysis applications executed locally	Add capability to compose processing and analysis applications in complex scientific pipelines	S1, S2, S3
Planetary Science	**RP1**	Automate image mosaicking	Computing	Mandatory	Using ISIS3	Currently technical-dependent on specialists in image-processing and planetary data.	Use machine-learning based features definition	S2
	RP2	Evaluate landing sites	Computing	Optional	Using SAGA-GIS	Currently technical-dependent on image-processing and statistical measurements.	Pre-define set of customization parameters, build user interface for batch processing	S1

19 Available values are: **Computing / Storage/ Cloud**
20 Available values are: **Mandatory / Convenient / Optional**

	RP3	Predict mineral resources	Computing	Mandatory	Using ArcGIS on template-based spectra	Currently missing ML/DL capabilities	Develop machine-learning workflow for mineral detection	S3
	RP4	Parallelise Rasdaman database manager	Storage/Cloud	Mandatory	Using serial implementation	Currently serial processing, restrict to one machine	Implement data distribution and (implement) load balance accordingly	S1
	RP5	Improve reproducibility	Cloud	Convenient	Using Docker containers, Jupyter notebooks	Currently based completely on documentation and custom scripts	Standardise software deployment, and documentation.	S2, S3

Table 5 Mapping of user requirements with co-design specification

4. Software Development Plan and Guidance

4.1. NEANIAS Space Services Architecture

The technical objectives of the WP4 Space activities are focused on the implementation of a common architecture for the provision of space services. The preliminary common architecture, depicted in Figure 9, is the result of integration among the available software components, datasets and infrastructures. The WP4 Software Development Plan has the purpose to get standalone elements that are available today and evolve them to the required TRL in order to allow the implementation of the Space Service Architecture and its porting to EOSC.

The preliminary architecture has been defined through both the decomposition of existing tools and applications into modular elements and the identification of all external dependencies that are mandatory to run the applications. All identified single elements have been reconnected applying these principles of:

- maximising the reuse of software components;
- use technology that are state-of-the-art;
- ensure the system portability on a cloud architecture;
- build a performant system able to exploit huge and complex datasets.

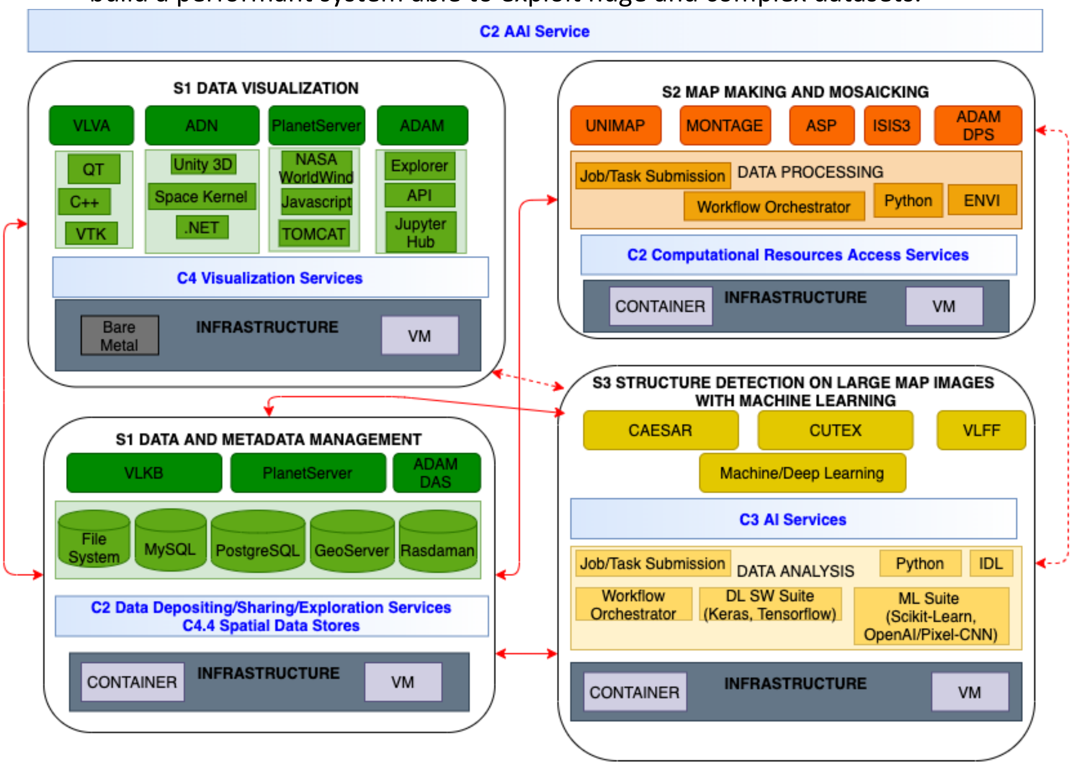

Figure 9 NEANIAS Space Services Architecture

In addition, another goal of the Space Service's preliminary architecture is to identify all technical requirements that must be considered to form a basis for both WP6 "Core Services" and WP7 "Service Delivery" system definition and design activities.

A common architecture as proposed is currently presented by grouping elements according to the envisaged space services as definition, stage decomposition and aggregation activities are still ongoing. However, the core elements that are needed to integrate and provide the space services have been identified and they are briefly described herein. The Space Services architecture diagram in Figure 9 also shows relationships (red arrows, dashed line means they are still TBC, i.e. to be confirmed) between groups, showing that components can consume services provided by components in other blocks. Finally, light blue boxes show preliminary identified Core Services that will be developed in WP6 supporting the Space Services.
The application level consists of applications that were available before NEANIAS (at TRL 6) and provide the space service to end-users through HMI interfaces. All technical activities are based on refactoring, repurposing and integrating upon these applications.
For each application, the development framework and the execution engine have been identified and are shown as independent blocks in the architecture schema. Examples of development framework are .NET and QT for the data visualisation applications or ENVI for S2 existing tools. In a similar way, external software libraries have been identified, some examples being the Space kernel for the ADN application and DL/ML software libraries (such as scikit-learn or TensorFlow) for the S3 machine learning services. For those services that can be provided as data processing or analysis tasks, against a user request, the infrastructure software of workflow orchestrator, job/task submission and resource cluster management have been introduced as new software components needed to build services above the existing processing application and algorithms implementation.
Data and metadata layer, being a core architecture part, is already at this stage separated from S1 service group as, surely, it will provide data access services to all other software components. Data and metadata layers consist of the available applications: VLKB, Planet Server and ADAM including all inner data stores that persist data, properly integrated on a common software layer providing API for consumers.
The lower layer is the infrastructure one that provides computation, storage and network resources through either bare metal or virtualised solutions, i.e. virtual machine (VM) and containers. The architecture schema shows which are the current infrastructure needs that could change during the process of deployment of the Space Services on EOSC. Finally, the architecture blocks will expose their services behind Application Programming Interfaces (API) in order to harmonise the services access and utilisation and improve their integrability within other systems.

4.2. Software Validation

The space services will be validated using the datasets reported in the Data Managements Plans prepared for Deliverable 1.5 [24] and are summarised in the following section.
For developing and validating the Space Service #1 (S1), we aim to employ the following datasets selected by the Astrophysics and Planetary Science communities.

DATASETS - S1					
id	**Name of Dataset**	**Type**	**Source**	**FAIR**	**Size** (GBs)
Astrophysics					
DS1-1	SGPS (HI)	Fits files (datacube)	Australian Telescope Compact Array + Parkes Single Dish	Y	4.4 GB
DS1-2	WISE (all-Bands)	Fits files (images)	Wide-field Infrared Survey Explorer (WISE)	Y	176 GB
DS1-3	CORNISH (5GHz)	Fits files (images)	Co-Ordinated Radio 'N' Infrared Survey for High-mass star formation	Y	84 GB
DS1-4	MIPSGAL (24µm)	Fits files (images)	Spitzer/MIPS Galactic Plane Survey	N	13 GB
DS1-5	Hi-GAL (all-bands)	Fits files (images)	Hershel infrared Galactic Plane Survey	Y	18 GB
DS1-6	FCRAO GRS	Fits files (datacube)	Boston University – FCRAO Galactic Ring Survey	N	11 GB
DS1-7	ThrUMMS (12CO 13CO C18O)	Fits files (datacube)	Three-mm Ultimate Mopra Milky Way Survey	N	35 GB
DS1-8	CGPS (HI)	Fits files (datacube)	Canadian Galactic Plane Survey	N	45 GB
DS1-9	Hipparcos Catalogue	Plain text file	ESA Hipparcos Catalogue	N	Less than 10 MB
DS1-10	Gaia Catalogue	Gbin files RDBMS	Gaia DPCT	Y	1.5TB 24TB
Planetary Science					
DS1-11	ESA MEX HRSC Nadir imagery	Geotiff	ESA PSA	Y	1+ TB
DS1-12	NASA MRO CRISM	PDS/ISIS cube/geotiff	NASA PDS	N	20 TB
DS1-13	NASA MRO HIRISE	PDS/ISIS cube/geotiff	NASA PDS	N	20 TB

Table 6 Datasets for validation of the S1 service

In particular, VisIVO, VLVA and VLKB services have been already tested on the datasets from DS1-1 to DS1-8. ADN software have been tested on the Hipparcos datasets DS1-9 and will be extended to use Gaia datasets DS1-10. Also, PlanetServer has already been tested on the datasets from DS1-11 to DS1-13 and run on in-house computing infrastructure.

For developing and validating the Space Service #2 (S2), we aim to employ the following datasets:

DATASETS - S2					
id	**Name of Dataset**	**Type**	**Source**	**FAIR**	**Size** (GBs)
		Astrophysics			
DS2-1	VLASS	Fits files	Very Large Array Telescope All-Sky Survey	Y	900 GB
DS2-2	GLIMPSE (8μ)	Fits files	Spitzer Space Telescope	Y	32 GB
DS2-3	MOPRA DR3	Fits files	Mopra Southern Galactic Plane CO Survey	Y	240 GB
DS2-4	SGPS (continuum 1.4GHz)	Fits files	ATKA + PARKES	N	25 MB
DS2-5	WISE (11μ, 22μ)	Fits files	Wide-field Infrared Survey Explorer (WISE)	Y	88 GB
DS2-6	Hi-GAL raw data	Fits files	INAF IAPS	N	2 TB
		Planetary Science			
DS2-7	ESA MEX HRSC DTM	PDS/ISIS cube/geotiff	ESA MEX (PSA)	Y	1+ TB
DS2-8	NASA MRO CTX	PDS/ISIS cube/geotiff	NASA MRO	Y	10 TB
DS2-9	NASA LRO LROC	PDS/ISIS cube/geotiff	NASA PDS	Y	10 B
DS2-10	NASA MRO LOLA + KAGUYA	PDS/ISIS cube/geotiff	NASA PDS / JAXA	Y	1 TB

Table 7 Datasets for validation of the S2 service

Finally, for developing and validating the Space Service #3 (S3), we will employ the following datasets:

DATASETS - S3					
id	**Name of Dataset**	**Type**	**Source**	**FAIR**	**Size (GBs)**
		Astrophysics			
DS3-1	SCORPIO-BAND1 0.9GHz	IMAGE RADIO CONTINUUM fits file	ASKAP EMU EARLY SCIENCE (Private Data)	N	1.5 GB
DS3-2	SCORPIO-BAND2 1.2GHz	IMAGE RADIO CONTINUUM fits file	ASKAP EMU EARLY SCIENCE (Private Data)	N	1.5 GB

DS3-3	SCORPIO-BAND3 1.6GHz	IMAGE RADIO CONTINUUM fits file	ASKAP EMU EARLY SCIENCE (Private Data)	N	1.5 GB
DS3-4	SKA – SIMULATED BAND 1.4 GHz	IMAGE RADIO CONTINUUM fits + ascii files	SKA DATA CHALLENGE	Y	4.5 GB
DS3-5	Catalogues of point sources (such as WISE, 2MASS)	Ascii files to use as DL training sets	Vizir	Y	1 GB
DS3-6	Hi-GAL (all bands)	IMAGES Fits file	Hi-Gal - Herschel Galactic Plane Survey	Y	18 GB
Planetary Science					
DS3-7	ESA MEX HRSC Nadir imagery	PDS/ISIS cube/geotiff	ESA PSA	Y	1+ TB
DS3-8	ESA MEX HRSC DTM	PDS/ISIS cube/geotiff	ESA PSA	Y	1+ TB
DS3-9	NASA MRO CTX	PDS/ISIS cube/geotiff	NASA PDS	Y	10 TB
DS3-10	PLANMAP maps	GeoTIFF/Geopackage	Planmap consortium	Y	5 TB
DS3-11	Mars Cave Database	Shapefile	NASA USGS	N	2 MB
DS3-12	Mars Global Digital Dune Database: MC–1	GeoTIFF/Shape file	NASA USGS	N	2 GB
DS3-13	Mars Global Digital Dune Database: MC2–MC29	GeoTIFF/Shape file	NASA USGS	N	2 GB
DS3-14	Mars Global Digital Dune Database: MC–30	GeoTIFF/Shape file	NASA USGS	N	3 GB

Table 8 Datasets for validation of the S3 service

The structure detection software for implementing S3 for the Astrophysics community is given by CUTEX, VLFF and CAESAR, and those have already been tested on the mentioned astrophysics datasets and will be validated with machine learning/deep learning techniques. For Planetary Science, machine learning techniques have not yet been tested on the datasets from DS3-7 to DS3-14 to perform: (i) image registration and/or (ii) features detection (e.g. to discover landing sites).

4.3. Software Releases

The Space Thematic services S1, S2 and S3 will be released in three main cycles. Release #1 will be due in M12 (Deliverable D4.3 and relative report in Deliverable 4.4). Release #2 will be due in M23 (Deliverable D4.5 and relative report in Deliverable 4.6). Final Release #3 will be due in M30 (Deliverable D4.7 and relative report in Deliverable 4.8). The following table summarize the pre-release, testing and production project months and estimated exact dates.

		From project start			Exact dates		
Deliverable	**Description**	**Production**	**Testing**	**Pre-release**	**Production**	**Testing**	**Pre-release**
D4.3	Space Thematic Services Release #1	12	11	10	31/10/20	30/09/20	31/08/20
D4.5	Space Thematic Services Release #2	23	22	21	30/09/21	31/08/21	31/07/21
D4.7	Space Thematic Services Release #3	30	29	28	30/04/22	31/03/22	28/02/22

Table 9 Software development plan of the S1, S2, S3 service releases including extimated dates for pre-releases, testing and production.

5. Conclusions and Future Steps

This deliverable presented the requirements of the NEANIAS WP4 “Space Research Services” including a detailed description of the user communities, datasets and products involved as well as the technological details of the tools in use and the foreseen architecture, development plan and validation for the three space services: S1-FAIR Data Management and Visualisation, S2 - Map making and mosaicking, and S3 - Structure detection with machine learning.

This document served to WP6 for the definition of the needed architecture, design principles and specifications on core services (deliverable D6.1) especially for the task T6.5 C3 - AI services implementation and the task T6.6 C4 - Visualisation services implementation. Furthermore, it has been shared with the User Board and to the WP7 for planning the services delivery and operation.

The deliverable will be continuously updated once new requirements come out (both from user and technological perspectives), with a new deliverable D4.2, due on M25, entitled “Space Research Services Report on requirements and specification” collecting all the final service specifications and requirements.

References

[1] F. Vitello et al., Vialactea Visual Analytics Tool for Star Formation Studies of the Galactic Plane, Publications of the Astronomical Society of the Pacific 130.990 (2018): 084503

[2] Molinaro, Marco, et al. "VIALACTEA knowledge base homogenizing access to Milky Way data." Software and Cyberinfrastructure for Astronomy IV. Vol. 9913. International Society for Optics and Photonics, 2016.

[3] L. Piazzo et al., *"UNIMAP: a generalized least-squares map maker for Herschel data"*, MNRAS, (2015), https://doi.org/10.1093/mnras/stu2453

[4] S. Riggi et al., *Automated detection of extended sources in radio maps: progress from the SCORPIO survey*, MNRAS 460, 1486 (2016), https://doi.org/10.1093/mnras/stw982

[5] S. Riggi, F. Vitello et al., *CAESAR source finder: recent developments and testing*, Publications of the Astronomical Society of Australia, 36, E037 (2019), doi:10.1017/pasa.2019.29

[6] S.Molinari et al., *Source extraction and photometry for the far-infrared and sub-millimeter continuum in the presence of complex backgrounds*, A&A 530, A133 (2011), doi:10.1051/0004-6361/201014752

[7] E. Schisano et al., "*The Hi-GAL catalogue of dusty filamentary features of Galactic Plane*", MNRAS, (2020), in press

[8] E. Schisano et al, "*The identification of filaments on far-infrared and submillimiter images: morphology, physical conditions and relation with star formation of filamentary structure*", ApJ, (2014), http://dx.doi.org/10.1088/0004-637X/791/1/27

[9] Anderson, J. A., Sides, S. C., Soltesz, D. L., Sucharski, T. L., & Becker, K. J. (2004). Modernization of the integrated software for imagers and spectrometers. In S. Mackwell & E. Stansbery (Eds.), Lunar and Planetary Science XXXV, Abstract2039. Houston (CD-ROM): Lunar and Planetary Institute.

[10] Beyer, Ross A., Oleg Alexandrov, and Scott McMichael. 2018. The Ames Stereo Pipeline: NASA's open source software for deriving and processing terrain data, Earth and Space Science, 5. https://doi.org/10.1029/2018EA000409.

[11] Gaddis, L., Anderson, J., Becker, K., Becker, T., Cook, D., Edwards, K., ... & Mathews, J. (1997, March). An overview of the integrated software for imaging spectrometers (ISIS). In Lunar and Planetary Science Conference (Vol. 28, p. 387).

[12] Marco Figuera, R., Pham Huu, B., Rossi, A. P., Minin, M., Flahaut, J., Halder, A. (2018) Online characterization of planetary surfaces: planetserver, an open-source analysis and visualisation tool, Planet. Space Sci. Volume 150, January 2018, Pages 141-156, DOI: 10.1016/j.pss.2017.09.007.

[13] Sault R.J., Teuben P.J., & Wright M.C.H., 1995. A retrospective view of Miriad, Astronomical Data Analysis Software and Systems IV, ed. R. Shaw, H.E. Payne, J.J.E. Hayes, ASP Conference Series, 77, 433-436.

[14] McMullin, J. P., Waters, B., Schiebel, D., Young, W., & Golap, K. 2007, Astronomical Data Analysis Software and Systems XVI (ASP Conf. Ser. 376), ed. R. A. Shaw, F. Hill, & D. J. Bell (San Francisco, CA: ASP), 127

[15] McCLure-Griffiths et al, 2006, 'The Southern Galactic Plane Survey: H I Observations and Analysis', ApJS, 158, 178

[16] Hoare, M. G.; Purcell, C. R.; Churchwell, E. B.; Diamond, P.; Cotton, W. D.; Chandler, C. J.; Smethurst, S.; Kurtz, S. E.; Mundy, L. G.; Dougherty, S. M.; Fender, R. P.; Fuller, G. A.; Jackson, J. M.; Garrington, S. T.; Gledhill, T. R.; Goldsmith, P. F.; Lumsden, S. L.; Marti, J.; Moore, T. J. T.; Muxlow, T. W. B.; Oudmaijer, R. D.; Pandian, J. D.; Paredes, J. M.; Shepherd, D. S.; Spencer, R. E.; Thompson, M. A.; Umana, G.; Urquhart, J. S.; Zijlstra, A. A., 2012, The Coordinated Radio and Infrared Survey for High-Mass Star Formation (The CORNISH Survey). I. Survey Design, PASP Vol 124, Pgs 939-995.

[17] Umana, G., Trigilio, C., Franzen, T.M.O., et al., 2015, SCORPIO: a deep survey of radio emission from the stellar life-cycle, MNRAS, 454, 902

[18] Carey, S. J. et al, 2009, MIPSGAL: A Survey of the Inner Galactic Plane at 24 and 70 µm, PASP, 121, 76

[19] Wright E. L., et al., 2010, The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance, AJ, 140, 1868

[20] Mainzer A., et al., 2011, Preliminary Results from NEOWISE: An Enhancement to the Wide-field Infrared Survey Explorer for Solar System Science, ApJ, 731, 53

[21] Molinari S., et al., 2016, Hi-GAL, the Herschel infrared Galactic Plane Survey: photometric maps and compact source catalogues. First data release for the inner Milky Way: +68° ≥ l ≥ -70°, A&A, 591, A149

[22] Jackson, J.M., et al., 2006, ApJS, 163, 145

[23] Peter Barnes; Erik Muller, 2016, "ThrUMMS Extended Observations: Spectral Line (3D)", https://doi.org/10.7910/DVN/CCCKQB, Harvard Dataverse, V1

[24] NEANIAS Deliverable 1.5, D1.5 Data management plan, report aligned to the H2020 suggested template and FAIR principles.

[25] CRISM mission website: https://mars.nasa.gov/mro/mission/instruments/crism/

[26] Becker K., Becker T, and Hare T., 2013, "MESSENGER MDIS Data User's Workshop 2013",https://pds-imaging.jpl.nasa.gov/documentation/MESSENGER_MDIS_WorkShop_LPSC2103.pdf

[27] Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3,** 160018 (2016). https://doi.org/10.1038/sdata.2016.18

List of acronyms

Acronym	Description
ADAM	Advanced geospatial Data Management platform
ADN	Astra Data Navigator
CRISM	Compact Reconnaissance Imaging Spectrometer for Mars
DL	Deep Learning
EOSC	European Open Science Cloud
EPN-TAP	EuroPlanet - Table Access Protocol
FAIR	Findable, Accessible, Interoperable, and Reusable
FITS	Flexible Image Transport System
IVOA	International Virtual Observatory Alliance
KPI	Key Performance Indicator
ML	Machine Learning
MRO	Mars Reconnaissance Orbiter
OGC	Open Geospatial Consortium
SKA	Square Kilometer Array
TAP	Table Access Protocol
VLFF	ViaLactea Filament Finder
VLKB	ViaLactea Knowledge Base
VLVA	ViaLactea Visual Analytics
WCS	Web Coverage Service
WFS	Web Features Service
WMS	Web Map Service
WPS	Web Processing Service

Appendix

UB Member	Contact	Brief CV
Chiara Marmo	chiara.marmo@inria.fr	Research Engineer at Inria (National French Institute for Computer Science Research). She has a PhD in Astronomy and a 15 years experience in data processing for Astronomy and Planetary Science. She worked in image mosaicking and visualisation, participating to the development of Open Source tools used in Astronomy (Astromatic suite) and Planetary and Geospatial research

		(GDAL). During her stay at GEOPS (Paris Saclay University- France) she was the coordinator of the participation of the laboratory at the Europlanet-RI 2020 project, working in particular on metadata standard definitions for improving interoperability between Astronomy and Planetary Science. She is now Community and Operation Manager at the scikit-learn Inria Foundation, which employs central contributors to the scikit-learn project, to support its community.
Lucia Marchetti	marchetti.lu@gmail.com	Senior Lecturer in the Department of Astronomy at the University of Cape Town. She obtained her MSc in Astronomy in 2008 and her Ph.D. in Astronomy in 2012 both from the University of Padova. Between 2012 and 2020 she has worked as Postdoc at the Open University in the UK and after at the University of Western Cape and University of Cape Town. Her research focuses on galaxy evolution and strong gravitational lensing studies. She co-authored >80 refereed scientific publications with > 4300 citations in total (15 publications have more than 100 citations each) and has an h-index of 35. She is involved in a number of international scientific consortia, co-leading scientific and technical working groups. She is a member of the European Commission Research Executive Agency and South Africa's Department of Science and Technology HELP project, whose goal is to complete the exploitation of the extragalactic surveys undertaken by the Herschel Space Observatory during its mission and foster scientific collaboration both within and between Europe and South Africa. She is a member of the International Astronomical Union (IAU), a fellow of the Royal Astronomical Society, and a member of the South African Institute of Physics (SAIP). She is also a member of the African Planetarium Association (APA) executive board and project manager of the IDIA visualisation Lab based at UCT.

User Board members selected to provide feedback for the Space Services Requirements