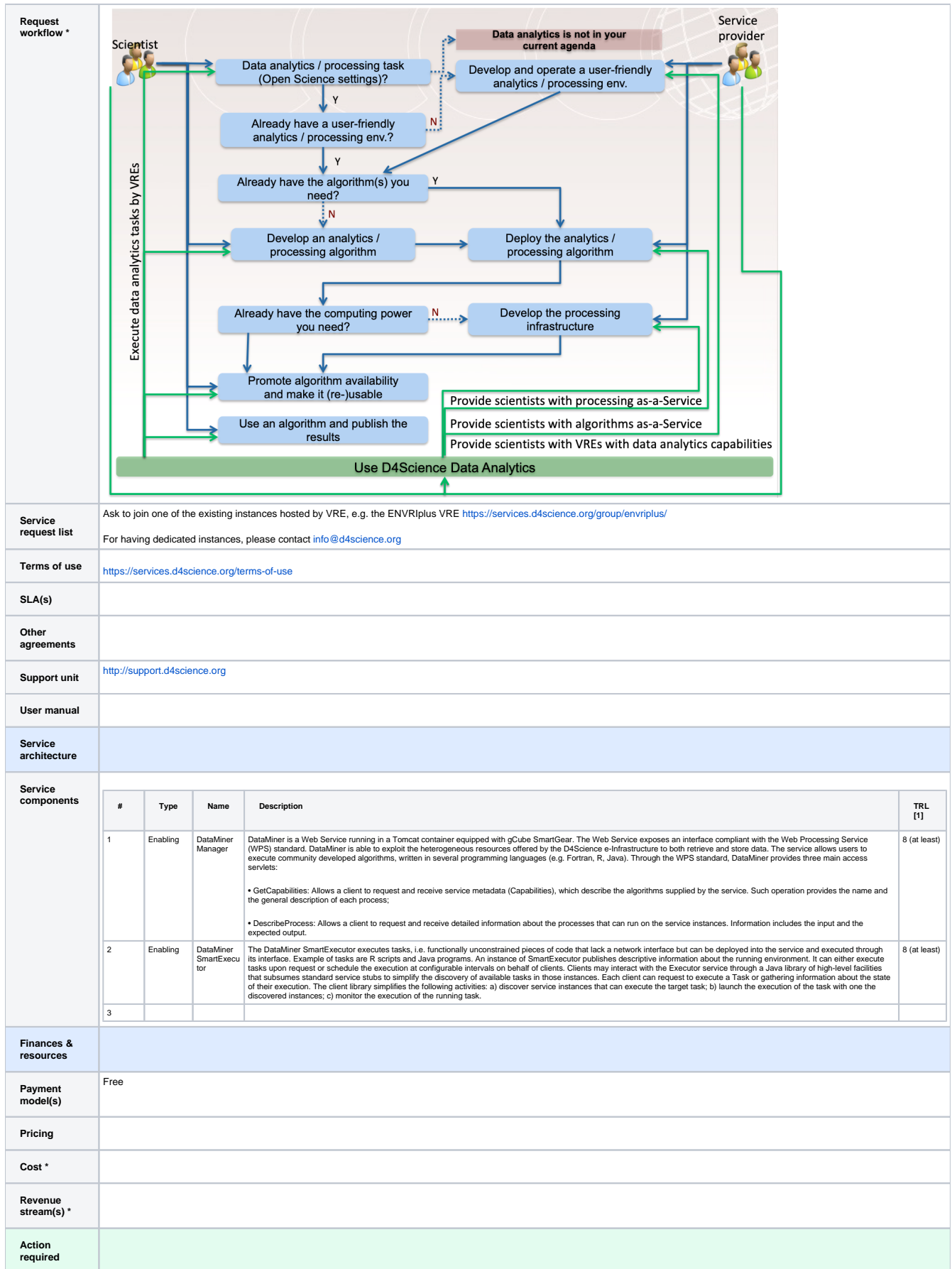


B2. gCube / D4Science DataMiner

The attributes marked with a * are confidential and should not be disclosed outside the service provider.

Service overview																									
Service name	gCube / D4Science DataMiner																								
Service area	Data processing and analytics																								
Service phase	Production																								
Service description	<p>This service offers a web-based workbench for data analytics compliant with Open Science practices. From the end user perspective, it offers a collaborative-oriented working environment where users:</p> <ul style="list-style-type: none">• can easily execute and monitor data analytics tasks by relying on a rich and open set of available methods either by using a dynamically generated web-based user- friendly GUI or by using a RESTful protocol based on the OGC WPS Standard;• can easily share & publish their analytics methods (e.g. implemented in R, Java, Python, etc.) to the workbench and make them exploitable by an automatically generated web-based GUI and by the OGC WPS protocol;• are provided with a “research object” describing every analytics task executed by the workbench enabling for repeatability, computational reproducibility, reuse, citation and provenance. The research object contains every input & output, an executable reference to the method as well as rich metadata including a PROV-O provenance record; <p>The data analytics framework is integrated with a shared workspace where the research objects resulting from the analytics tasks are automatically stored together with rich metadata. Objects in the workspace can be shared with coworkers as well as published by a catalogue with a license governing their uses. Moreover, the framework is conceived to operate in the context of one or more Virtual Research Environments, i.e. it is actually made available by a dedicated working environment offering (besides the framework and the workspace) additional services including those for managing users, creating communities, and supporting communication and collaboration among VRE members.</p> <p>The data analytics framework is conceived to give access to two typologies of resource:</p> <ul style="list-style-type: none">• a distributed, open & heterogeneous computing infrastructure for the real execution of the analytics tasks. This distributed computing infrastructure is capable to exploit resources from the EGI infrastructure.• the pool of methods integrated in the platform, i.e. each method integrated in the framework is made available as-a- Service to other users according to the specific policy; <p>More details on this framework is available at https://wiki.gcube-system.org/gcube/DataMiner_Manager</p>																								
Customer group	Any Research Performing Organization willing to provide its scientists with an Open Science compliant data analytics platform.																								
User group	<p>The service is not tailored to serve the needs of a specific community. Rather, it is community agnostic and highly and easily customizable thus to serve the needs of a given community. Customization is achieved by configuring the instance serving a certain Virtual Research Environment with (a) the set of methods to be made available as-a-Service and (b) the resources forming the distributed computing infrastructure dedicated to execute the analytics tasks.</p> <p>Up to now it has been and is successfully used by a quite rich array of diverse communities, namely those associated with the supported projects., e.g., i-Marine (fisheries and marine biodiversity scientists), BlueBRIDGE (fisheries and aquaculture scientists, educators & SMEs), SoBigData.eu (social mining scientists), ENVRI+ (environmental scientists), AGINFRA+ (agriculture scientists), EGIP (geothermal scientists).</p>																								
Value	<p>The analytics platform is conceived to serve the needs of scientists (in particular, those belonging to the so called long-tail of science) by providing them with an easy to use working environment (nothing need to be installed on users' machine). It is conceived to hide the technicalities related with the execution of tasks by relying on distributed computing infrastructures. Moreover, it is conceived to be exploitable by third-party software/applications, e.g. R-Studio, Q-GIS, or any workflow management system or application capable to interface with a RESTful service.</p> <p>Worth highlighting that the platform is Open Science “compliant” (e.g., every method is “published” and citable, every task leads to a research object) and Virtual Research Environment friendly, i.e. it is conceived to be customizable with respect to the methods to offer in a given application context as well as it is conceived to benefit from a collaborative environment for sharing artefacts and comments.</p> <p>Its characteristics make it particularly suitable to serve typical scientific contexts of the long tail of science.</p>																								
Tagline	Open, user friendly and extensible data analytics platform ready for Open Science and VREs.																								
Features																									
Service options	<table><tr><th>Option</th><th>Name</th><th>Description</th><th>Attributes</th><th></th></tr><tr><td>1</td><td></td><td></td><td></td><td></td></tr><tr><td>2</td><td></td><td></td><td></td><td></td></tr><tr><td>3</td><td></td><td></td><td></td><td></td></tr></table>					Option	Name	Description	Attributes		1					2					3				
Option	Name	Description	Attributes																						
1																									
2																									
3																									
Access policies	Policy-based and Wide-use. D4Science operates a number of instances of this																								
Service management information																									
Service owner *	D4Science.org																								
Contact (internal) *	info@d4science.org																								
Contact (public)	info@d4science.org																								



[1] Technology Readiness Levels (TRL) are a method of estimating technology maturity of components during the acquisition process. For non-technical components, you can specify "n/a". For technical components, you can select them based on the following definition from the EC:

- TRL 1 – basic principles observed

- **TRL 2** – technology concept formulated
- **TRL 3** – experimental proof of concept
- **TRL 4** – technology validated in lab
- **TRL 5** – technology validated in relevant environment (industrially relevant environment in the case of key enabling technologies)
- **TRL 6** – technology demonstrated in relevant environment (industrially relevant environment in the case of key enabling technologies)
- **TRL 7** – system prototype demonstration in operational environment
- **TRL 8** – system complete and qualified
- **TRL 9** – actual system proven in operational environment (competitive manufacturing in the case of key enabling technologies)