EGI-Engage

Report on the installed LifeWatch applications and their usage record

D6.18

Date	28 February 2017
Activity	WP6
Lead Partner	CSIC
Document Status	FINAL
Document Link	https://documents.egi.eu/document/3022

Abstract

The LifeWatch EGI-Engage Competence Centre was proposed with a clear mission: capture and address the requirements of Biodiversity and Ecosystems research communities for e-infrastructure services supporting the different applications, and, when adequate, promote the usage of EGI e-infrastructure. In this report we summarize the applications and services that have been considered for integration by the Competence Centre, describe how they can be implemented, following the service oriented architecture proposed for LifeWatch, analyse the status of their integration, and finally provide the corresponding information about their interest, current usage and future potential.

The EGI-Engage project is co-funded by the European Union (EU) Horizon 2020 program under Grant number 654142

COPYRIGHT NOTICE

DELIVERY SLIP

	Name	***Partner/Activity***	***Date***
From:	Jesus Marco	CSIC/SA2	16 Feb 2017
Moderated by:	Małgorzata Krakowian	EGI Foundation/WP1	
Reviewed by	Enol Fernandez Yin Chen	EGI Foundation/WP4 EGI Foundation/WP6	20/02/2017 21/02/2017
Approved by:	AMB and PMB		7/03/2017

DOCUMENT LOG

Issue	***Date***	***Comment***	***Author/Partner***
v0.1	23/07/2016	First draft	J. Marco/CSIC
v0.2	23/10/2016	Scheme to report on services	J. Marco, A. Lopez, E. Fernández, F. Aguilar (CSIC)
v0.3	6/12/2016	Draft including an example on Ecological Observatories and Models (ALGAEBLOOM)	J. Marco, F. Aguilar (CSIC)
v0.4	22/12/2016	First Inputs from different case studies included in the appendix	All LW-CC teams (see names in appendix)
v0.5	15/01/2017	Case studies detailed in the appendix	
v0.6	27/01/2017	Detailed information about services collected, classified, and revised	F. Aguilar, J.Marco /CSIC
v1.0	16/02/2017	Version ready for internal review	J. Marco/CSIC
v1.1	28/02/2017	Version including reviewers' suggestions	J. Marco/CSIC
FINAL		FINAL version	J. Marco/CSIC

TERMINOLOGY

A complete project glossary is provided at the following page: http://www.egi.eu/about/glossary/

Contents

Executive summary

The goal of the LifeWatch EGI-Engage Competence Centre (LW-EGI-CC) was to capture and address the requirements of Biodiversity and Ecosystems research communities.
Our final analysis includes 16 services in production addressing clear requirements in biodiversity and ecosystem research, and several of them are currently integrated and are using EGI FedCloud resources. These services are grouped under the four different categories originally considered in the Competence Centre proposal: Observatories, Workflows, Virtual Labs and Citizen Science support. Their interest expands to other research areas, not only those closer on specific functionalities, like for example Genetics or Earth Observation, but also on generic techniques, and in particular analytics: from support on HPC resources for R or Python, to deep learning using GPUs. Many of the services can be considered interesting for supporting innovation in the field of Big Data. The use of open source components and of data management standards makes these services even more suitable for further exploitation.

This analysis comes in a key moment for LifeWatch: the launching of the ERIC provides the perfect timing to better integrate and consolidate all these services that have been developed and are supported by 20 centres in 6 European countries into the European Open Science Cloud. This integration exploits the close connection that already exists with EGI, through this Competence Centre, and with projects such as INDIGO-DataCloud, using its solutions in complex Case Studies.

If the current experience of usage and impact, mainly based on local level support, is extended to a wider EU community, these services could provide an added value to the ensemble of e-infrastructure resources that will be available in the framework of the European Open Science Cloud. This would happen for example, with the transfer of a genomic pipeline workflow, such as TRUFA, which is now accessed by more than 200 users worldwide in the field of biodiversity, to applications in the health area; or think about the possibilities of a service oriented to citizen science like Natusfera that could easily enable the setup of projects where users can upload and discuss their observations, to reach more than 5.000 users, and support Latin America communities.

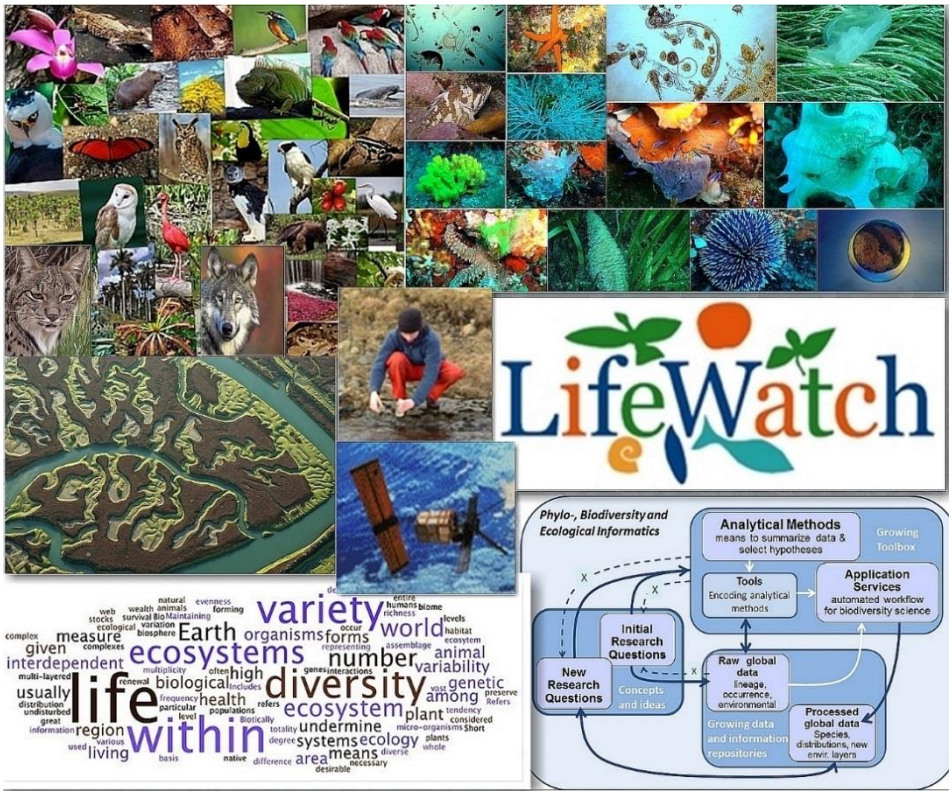
These examples of “visible” services for wider re-use, are balanced by those that are very mature and well known to the biodiversity community, such as the ones provided by GBIF Spain and Portugal, or by the advanced **LifeWatch Virtual Research Environments, launched on top of FedCloud resources since 2015, and exploited by first line research groups and centres that provide for example the references on the quality of our marine environment in Europe.**

We foresee an integration and implementation plan coordinated from LifeWatch ERIC in close contact with the next EINFRA12(a) project[1] management; our previous common experience, in particular in the EGI-Engage Competence Centre, is the best guarantee of our responsiveness and fulfilment of commitments, through an adequate internal organization, that will be reinforced with the ERIC launch.

[1] Identified as EOSC-Hub at the time of writing this report.

1. Introduction

LifeWatch*[2]*, *the e-Science European Research Infrastructure for Biodiversity and Ecosystem Research, is a distributed Research e-Infrastructure to* ***advance biodiversity research and to address the big environmental challenges*** *and support knowledge-based strategic solutions to environmental preservation. This mission is achieved by providing access to a multitude of data sets, services and tools enabling the construction and operation of Virtual Research Environments.*



LifeWatch is included in the 2016 **ESFRI** roadmap, having entered into operational phase in 2016, and legally becoming an ERIC at the beginning of 2017. The estimated costs are 66 M€ for capital value, and 10 M€/year for operation.

As an e-Infrastructure of distributed nature, LifeWatch is composed[3] by Common Facilities, located in Spain (Statutory Seat and the ICT e-Infrastructure Technical Offices), Italy (Service Centre) and The Netherlands (Virtual Laboratories and Innovations Centre):

The **Statutory Seat and the ICT e-Infrastructure Technical Offices** will jointly assist to the coordination and management of the day-to-day institutional relationships, administrative, legal,

[2] http://www.lifewatch.eu

[3] https://ec.europa.eu/research/infrastructures/pdf/esfri/esfri_roadmap/esfri_roadmap_2016_adopted.pdf

and financial issues. Those include, among others, technology transfer, procurement and IPR matters, and the formal agreements with all the external data and e-Services suppliers, and the Service Legal Agreements (SLA) with local, regional, national and international entities, including decision makers and environmental managers. Also, they will coordinate and manage the ICT e-Infrastructure distributed construction, maintenance and deployment operations, including coordination of the design and implementation of e-Services demanded by the Service Centre, the Virtual Laboratories and Innovations Centre, as well as other Distributed Facilities. The **Service Centre** will provide the interface with the Biodiversity Scientific Community, identify the needs of the multiple user groups from different domains and areas of interest and coordinate the development and operation of those Services related. Also, they will assist in deploying the Services provided by the LifeWatch Research Infrastructure, including those enabling discovery, visualization, and download of data and applications for analysis, synthesis and modelling of Scientific topics. Thus the Service Centre will identify new data resources, incorporate vocabularies, semantics and Services to aggregate larger typologies of data. It will also provide the optimization of the access and use of Service Centre facilities as a whole, and offer web-based tools to facilitate Social Networking and Social Learning (including e-Learning). Finally, it will promote the awareness of LifeWatch for users and general public, and the enhancing the visibility of LifeWatch scientific outcomes, by publicizing and disseminating them. The **Virtual Laboratories and Innovations Centre** will coordinate and manage the requirements and needs analysis, design and implementation of the scientific case studies and productions of the LifeWatch Virtual Laboratories. These e-Labs will be implemented and deployed through the LifeWatch ICT distributed e-Infrastructure facilities, and made accessible through the Service Centre to the Biodiversity Scientific Community. This procedure will guarantee the overall coherence of the Research Infrastructure by promoting synergies in regards to the semantic interoperability among data, services and their final users.

Distributed Facilities - Member countries of the LifeWatch ERIC and scientific networks are encouraged to establish LifeWatch Centres to serve specialized facilities in the framework of the LifeWatch services, in accordance with it overall architectural scheme. They are located in another four member countries (Belgium, Greece, Portugal, Romania and Slovenia); the list of participant countries extends to 14, including France, Finland, Hungary, Norway, Slovakia and Sweden.

LifeWatch allows its users to enter new research areas supported by its **e-Infrastructure**, building capacity to foster new opportunities for large-scale scientific development; to enable accelerated data capture with new technologies; to support knowledge based decision making for the management of biodiversity and ecosystems; and to support training programs.

Connecting LIFEWATCH and EGI

Since the design phase LifeWatch architecture has been based in a Service Oriented Approach, exploiting the possibilities offered by common e-infrastructures, and in particular exploring the option of cloud-based services, as quoted already in the LW statement for H2020 in 2013.

As shown in figure 1, LifeWatch integrates different data resources, some of them more specific to biodiversity and ecosystems research, like collections or measurements and observations series from ecological observatories, others of wider scope, like genetic information, satellite images or meteorological measurements.

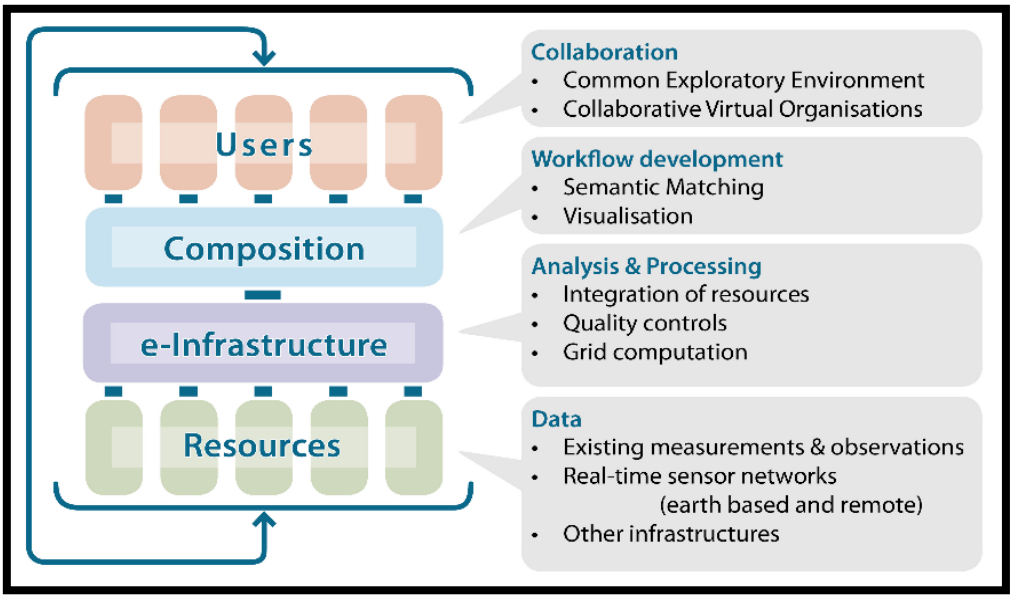


Figure 1: LifeWatch basic architecture scheme

EGI e-Infrastructure, and in particular EGI FedCloud resources, was considered one of the best options to provide the e-infrastructure layer required, where analysis and processing tools can be deployed and executed. The **LifeWatch Competence Centre under the EGI-Engage** project addressed the adoption and exploitation of the EGI infrastructure by the LifeWatch user community, in particular by deploying basic tools required to support data management, data processing and modelling for Ecological Observatories, considering the services required to support workflows oriented to the deployment of Virtual Labs for LifeWatch, and exploring the possibilities to offer a better support to the direct participation of citizens in LifeWatch contributing observation records, in particular those including images uploading and processing.

The collaboration established between EGI and LifeWatch through this Competence Centre along the last two years has been quite successful, with very active participation of different LW centres and also of several NGIs related to the LW participant countries, resulting in a rich catalogue of potential solutions, presented below, many of them already integrated and in operation using EGI FedCloud resources.

Definition of Services

The service portfolio presented in what follows represents the expertise in provisioning services, tools and platforms as introduced before, that is fully described in the different services templates below. The added value of this consortium is the joint expertise both on the specific scientific research area, and on the integration with the required e-infrastructures. The expertise represented by LIFEWATCH ERIC is based on the **contrasted and complementary experience of close to 20 different research groups in six EU countries** that will continue their collaboration in the next years under the LIFEWATCH ERIC umbrella.

The "Scientific and Technical description of LifeWatch ERIC" document[4], includes the following figure *showing some examples of data, modelling and analysis capabilities based on the expected e-Services provided by LifeWatch distributed e-Infrastructure.*

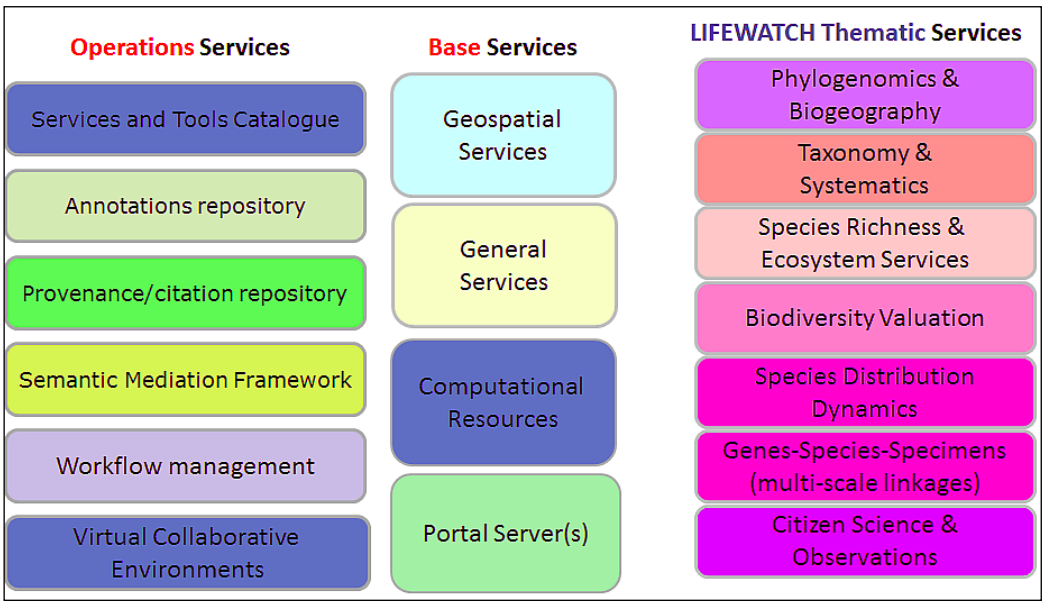


Figure 2: Examples of main services provided by LifeWatch e-Infrastructure

This initial scheme is in evolution to match the **realistic requirements of the researchers** in their different projects and initiatives, and also take into account new research and technical advances. The list of services analysed covers many of those requirements and incorporates also some of those advances, and one of our key targets is to consolidate our service organization scheme taking also into account the connection and integration with the basic e-infrastructure services. The collaboration with EGI initiative is very important for LifeWatch to achieve this objective, and at the same time our demanding requirements and the **richness of our stakeholders, including public administration, consultancy companies, many of them SMEs, and citizen scientists, provides a real challenge to our future objectives**.

The following table summarizes the list of services considered within the Competence Centre:

[4] http://lifewatch.eu/-/m/851/Scientific%20and%20Technical%20description%20of%20LifeWatch%20ERIC.pdf

Service Name	Main area	Main teams involved	Status*
1-Collaborative platform for observatories	**Observatories** Data Management	LW-Be, LW-Gr [VLIZ (Be), HCMR (Gr)]	**Integrated in production using FedCloud resources**
2-Modelling Water Masses	**Observatories** Modelling	JRU LW.ES, NGI-ES, NGI-IT [IFCA(Sp), Ecohydros(SME) (Sp), CITIC (Sp), INFN(It)]	**In production using partially FedCloud resources**
3-Data Services	**Observatories** Data Management	LW-Gr [HCMR (Gr)]	In production at HCMR
4-GBIF data access biogeographic context	**Observatories** Data Collections	JRU LW.ES, LW-PT, NGI-ES, NGI-PT [GBIF Spain and GBIF Portugal, LIP(Pt), IFCA(Sp)]	**In production using partially FedCloud resources**
5-Citizen Science Services	**Citizen Science** Data Processing	JRU LW.ES, NGI-ES [BIFI(sp), CREAF (Sp), GBIF node (Sp), IFCA (Sp), U.Granada (Sp)], U.Cordoba (Sp)].	**Integrated and in production using partially FedCloud resources**
6-Image Classification Deep Learning Tools	**Citizen Science** Data Analysis	JRU LW.ES, NGI-ES [BIFI (Sp), IFCA (Sp)]	**Developed, integrated and in production using FedCloud resources**
7-Genetic Services	**Virtual Lab**	LW-Gr [HCMR (Gr)]	In production at HCMR
8-MiroCT	**Virtual Lab**	LW-GR [HCMR(Gr)]	In production at HCMR
9-R Services	**Virtual Lab** Analysis	LW-Gr, LW-Be, JRU LW.ES [HCMR (Gr), VLIZ (Be), IFCA (Sp)]	In production at HCMR, **Integrated and in production using partially FedCloud resources**
10-Semantic Tools	**Virtual Lab** Data	LW-IT [UniSalento (It)/INFN]	In production at UniSalento
11-Phytoplankton VRE	**Virtual Lab** Data Analysis	LW-IT [UniSalento (It)/INFN]	In production at UniSalento
12-Ecological Data analysis platform	**Virtual Lab** Data Analysis	LW-IT [UniSalento (It)/INFN]	In production at UniSalento
13-Digital Knowledge Preservation Framework	**Virtual Lab** Data Management Analysis	JRU LW.ES [IFCA (Sp), U.Sevilla(Sp), CITIC(Sp)]	**In production using partially FedCloud resources**
14-Remote Monitoring and Smart Sensing	**Virtual Lab** Data Analysis	JRU LW.ES [IFCA (Sp), U. Sevilla (Sp), Andalucia SmartCity (Sp)]	In production at JRU-LW-ES
15-TRUFA	**Workflows**	JRU LW.ES [IFCA (Sp)]	In production at IFCA
16 -Declic	**Workflows**	NGI-FR, +NGI-ES [PGTB(Fr), UPV(Sp), France-Grilles]	In production at PGTB

All the services included in the previous table have been analysed in the EGI-Engage LifeWatch Competence Centre meetings regarding their interest as services in the LifeWatch context, however not all of them are at the same level of integration with EGI FedCloud infrastructure, as indicated in the **status** column:

-Developed: the service has been developed as part of the EGI-Engage LW CC effort.

-Integrated: the service has been integrated to use EGI FedCloud resources, as part of the CC effort.

-In production using (partially) FedCloud resources: the service is available, it uses EGI FedCloud resources, and it is considered of interest for LifeWatch researchers. The tag **partially** applies to complex services combining different resources, some of them external to FedCloud.

-In production (at *"LW Distributed Centre")*: the service is in production using local resources at a LW centre (like UniSalento or HCMR), and it has not been directly funded by the project, but it is considered of interest for LifeWatch researchers, and **its integration as a thematic service under the European Open Science Cloud has been proposed by the CC, within a 24 months framework.**

Regarding how these services can globally contribute to the challenges we need to address in our biodiversity field and in general, and although the very specific information is included in the different templates included below, it can be noted that:

- several of the services are dedicated to analytics, from those supporting the use of daily tools like R or python accessing HPC resources, to those exploiting new techniques like deep learning using latest generation of GPUs; as they are offered through web interfaces they can provide "easy user entry points", and not only to LifeWatch users but to a very wide range of researchers interested in these tools for analytics.

- most if not all the services are based on open source components, as this is a required condition in the LifeWatch initiative. Moreover, the biodiversity and environmental community is well known by playing a pro-active role in the development, promotion and adoption of standards, and not only in its own areas. As relevant examples in the services proposed, one can cite the use of standards for access and exchange of geo-based information, the promotion of the EML (Ecological Metadata Language), the use of taxonomic standards in GBIF, and more in general the promotion of RDA outcomes and more explicitly a FAIR+R approach, as proposed for example in the platform for knowledge preservation service. At a more basic level, the services are deployed or proposed to be deployed into the Cloud infrastructure adopting open standards, like those supported by the INDIGO-DataCloud initiative, and implemented in EGI FedCloud infrastructure.

A clear advantage of this approach, and basic to LifeWatch given the key role to be played by other stakeholders like public administrations and external companies developing or exploiting solutions, is the possibility for interoperation. LifeWatch has a very wide, and we hope also correct, vision on this topic, and we have made a bet not only to be able to interoperate at the e-infrastructure level, for example by considering the use of hybrid cloud resources, as currently

explored in projects like INDIGO-DataCloud, but also to be able to cover this interaction along the full application lifecycle, following an *Agile* and *DevOps* approach, that enables the tight collaboration between different actors from the start (requirements, design) to the end (deployment, performance, validation, support). This global vision on open development and implementation of services is in our opinion one of the differential features of our services, and also one of the best guarantees of the sustainability of the "production-level" of our services in the future.

Exploiting the possibilities offered by the cloud framework.

Regarding the explicit list of advantages and possibilities offered by the use of a cloud framework, we also try to summarize the current status:

-Most of the services are designed taking into account the integration with federated AAI, compute, storage and data management services provided by EGI. The main links are detailed in the summary table presented, but further details are provided in the detailed templates. As an example, many of the services are running on EGI FedCloud virtual machines, the LifeWatch VO is also operational since two years ago, several applications are using container solutions or exploiting the access to HPC resources. Moreover, LifeWatch is committed to support a common AAI solution developed in the EINFRA framework and that is the main reason for its involvement in the AARC2 initiative as well.

The list of services covers a wide range of digital capabilities (from data acquisition, storage, management, processing, analytics and visualization) on different dataset types, some as simple but interesting as historical records of temperature or radiation at ecosystems, others as complex as high resolution satellite images in different bands from last generation Earth Observation programs.

Another target within LifeWatch is the support of the offered services by an adequate management, help desks, and technical documentation. This is an on-going effort, that we expect to further expand within our next project, focussing mainly on the needs from communities new to the field that want to explore and use the services for multidisciplinary projects.

The services have shown, through their use, their relevance to international research collaborations, and are open, under the corresponding policies, to international user groups. As a relevant example, the TRUFA genomic pipeline, has provided more than 2 million hours of computing time in an HPC system in the last months to more than 200 researchers worldwide, an impact also reflected in the publications that have benefited of this usage.

As already indicated all the services are implemented on standard-based open architectures and technologies, and, even more, LifeWatch promotes the development of open source solutions, and the integration of the development under a DevOps orientation promoting also an Agile approach. We consider this as an important step required by services in production, an approach already adopted in many other initiatives. The possibility of a transparent test and validation of the services and their implicit dependencies on other (mainly open) components is key to this.

Along this line, the detailed templates specify the relevant set of user groups that have used the services described, and in some cases also provide evidence of the positive results. Notice however the difference between a production level for some of the services in operation for many years, and a production level for the same service after integration with the EINFRA services offered by EGI.

Most if not all the services presented already benefit from a high level of exploitation: services like GBIF.ES or GBIF.PT are serving the relevant communities with a high demand; relatively new applications for Citizen Science like Natusfera have already collected more than 5000 images and the user community has more than 1000 individuals; the marine VRE is in daily usage by the researchers at VLIZ and HCMR centres to assist in the collection and processing of the observations providing the monitoring measurements of the different marine areas in Europe, etc. etc. Regarding the innovation potential, and in particular in relation to the collaboration with the industry, we can point to the implementation of deep learning techniques for image recognition using large training datasets, or to the parameter sweeping to find the best modelling of algae bloom in a water reservoir in collaboration with an SME.

2. Potential impact and current usage

Analysis of potential impact of LifeWatch services

Our exploitation plan is based on the fact that the services offered are becoming part of the LifeWatch ERIC services, and so **supported by the detailed technical and financial plan officially approved for LifeWatch as an ESFRI**.

As described in the official ESFRI roadmap for 2016[5], LifeWatch initiative joins 7 EU countries as members and another 7 as participants, and represents an investment of >65M€ and an annual operation budget of 10M€.

The biodiversity and ecosystem research community that LifeWatch services target is very wide. First of all, it includes researchers in biodiversity and ecosystems research, mainly biologists and environmental scientists, but also in other technical related fields, in different research centres and universities, and also in museums and management of natural reserves. LifeWatch services will be available to these researchers, and an estimation of the potential number of users exceeds 5.000 EU researchers and more than 10.000 worldwide. This community publishes more than 10.000 papers/year only on the specific area of "biodiversity" (according to the Web of Science[6]). A second group of users includes environmental agencies (at European, national, regional and local level), and includes also consultancy and engineering companies, involved in management actions. And another very large and increasing group of final users are associations and citizen scientists.

The orders of the potential number of users for the next 5 years of LifeWatch services can be estimated to be: o(3.000) for basic researchers, o(500) for researchers involved in management of biodiversity and natural resources, o(10.000) for citizen scientists.

Current and expected use

The **current number of users of the services presented** are 1.000 for basic researchers (notice that for example GBIF.ES and GBIF.PT provide data service to their national communities, or that one of the popular services, TRUFA, has more than 200 active users submitting pipelines in the last year), o(200) for researchers involved in management (this number includes those participating in different official monitoring activities and also specific projects, in particular LIFE+ projects), and o(2.000) for citizen scientists (notice that >1000 were already active in the last year in the Natusfera project focused mainly in Spain).

[5] http://www.esfri.eu/roadmap-2016
[6] Web of Science, see wokinfo.com

The current usage statistics and a fair estimation of the potential evolution, both provided by the groups responsible of the different services, is provided in the following table:

Service Name	Main teams involved	Current # users	Potential #users 2020
1-Collaborative platform for observatories	LW-Be, LW-Gr [VLIZ (Be), HCMR (Gr)]	**80**	**>500**
2-Modelling Water Masses	JRU LW.ES, NGI-ES, NGI-IT [IFCA(Sp), Ecohydros(SME) (Sp), CITIC (Sp), INFN(It)]	30	**>100**
3-Data Services	LW-Gr [HCMR (Gr)]	50	>1000
4-GBIF data access under biogeographic context[7]	JRU LW.ES, LW-PT, NGI-ES, NGI-PT [GBIF Spain and GBIF Portugal, LIP(Pt), IFCA(Sp)]	**>10.000**	**>20.000**
5-Citizen Science Services	JRU LW.ES, NGI-ES [BIFI(sp), CREAF (Sp), GBIF node (Sp), IFCA (Sp), U.Granada (Sp)], U.Cordoba (Sp)].	>1000	**>5000**
6-Image Classification Deep Learning Tools	JRU LW.ES, NGI-ES [BIFI (Sp), IFCA (Sp)]	20	**>1000**
7-Genetic Services	LW-Gr [HCMR (Gr)]	30	**>500**
8-MiroCT	LW-GR [HCMR(Gr)]	20	**50**
9-R Services	LW-Gr, LW-Be, JRU LW.ES [HCMR (Gr), VLIZ (Be), IFCA (Sp)]	50	**>1000**
10-Semantic Tools	LW-IT [UniSalento (It)/INFN]	90	**>1000**
11-Phytoplankton VRE	LW-IT [UniSalento (It)/INFN]	70	**>500**
12-Ecological Data analysis platform	LW-IT [UniSalento (It)/INFN]	90	**>1000**
13-Digital Knowledge Preservation Framework	JRU LW.ES [IFCA (Sp), U.Sevilla(Sp), CITIC(Sp)]	30	**>50**
14-Remote Monitoring and Smart Sensing	JRU LW.ES [IFCA (Sp), U. Sevilla (Sp), Andalucia SmartCity (Sp)]	50	**>200**
15-TRUFA	JRU LW.ES [IFCA (Sp)]	**>200**	**>500**

Note: numbers in bold refer to users at international level.

[7] Detailed statistics: datos.gbif.es (Fev2016-Jan2017): **9761 users**, 16875 visits, datos.GBIF.pt (3 months, Nov2016-Jan2017): **713 users**, 1154 visits.

Appendix: Description of the services and their integration status with EGI e-Infrastructure

1-Collaborative platform for observatories

Service overview	
Thematic Service name	**Collaborative Platform for Observatory and sensor data analysis**
Service description	This service provides possibilities to collaborate with a team of researchers on the data analysis of the data generated by an operational marine observatory. The data flows through predefined pathways to a secure data store. A data explorer that provides a series of interactive user interfaces to the scientist is available for preliminary analysis of these data. This data explorer is supported by a performant stack of data storage, GIS and analysis systems. The exploratory analysis can be extended in a secure coding environment where researchers have direct access to the data and where code sharing, publication and versioning is facilitated. The service already provides analytical services for a broad set of sample and sensor based data and can be extended to additional types of sensors.
Service provider	*LIFEWATCH ERIC* *SERVICE PROVIDED BY VLIZ*
Service catalogue	LIFEWATCH (MARINE) ANALYTICAL SERVICES
Value	Easy explorative visualization and mapping, highly interactive, performant analytical environment. Shared coding and development, integrated data and processing.
Current TLR level, acceptance criteria and validation/ verification results	***TRL8*** *In production after being tested by both internal and external users* *Used by Belgium observatory data users (users@lifewatch.be): Phd students; researchers from research institutes and labs; use during analysis workshops.* *Needs further integration and upscaling* *Components online: http://www.lifewatch.be/en/lifewatch-data-explorer, http://rstudio.lifewatch.be/auth-sign-in,*

	Components presented at various fora. For example: -VLIZ Marine Scientist Day. Brugge, Belgium, 12 February 2016 *Deneudt, K.; Maes, P.; Vanhoorne, B.; Hernandez, F. (2016). Building an online and interactive scientific data explorer for LifeWatch observatory data, in: Mees, J. et al. (Ed.) Book of abstracts – VLIZ Marine Scientist Day. Brugge, Belgium, 12 February 2016. VLIZ Special Publication, 75: pp. 47,* -Lifewatch technical meeting, Crete, 3-5 June 2014 *https://www.lifewatchgreece.eu/sites/default/files/pdf_files/Collaborative-scientific-platforms.pdf*
Access policy	Policy based. Data comes into the public domain after moratorium period.
Terms of use	After moratorium data is published as a cc-by data publication. Upon publication of result LifeWatch needs to be referenced.
User groups and scientific disciplines served	Biodiversity, Ecology, Ecosystem Research, Ocean...
Service business model	The marine observatory is supported by Flemish government in the framework of the Flemish contribution to Lifewatch.

Service components

Name of component	Functional description, applicable standards and needed resource capacity (if applicable) e.g. CPU Time, storage capacity etc.	Provider If already appointed
Observatory Data	Data from Lifewatch observatory Belgium	VLIZ
Data store: MongoDB	File based sensor data storage	VLIZ
Data store: PostGress database + GEOSERVER	Spatial data storage and output	VLIZ

	R server	Analytical component for scripting processes	VLIZ
	RShiny Server apps	Analysis and visualization interface based on forms.	VLIZ
	Repositories	Repositories for scripts and versioning.	SVN, Github
Service integration with generic e-Infrastructures			
Integration activity and concerned service components	EGI FEDCLOUD, LIFEWATCH e-INFRA Further integration of components and upscaling performance needed		
List of requested service components	EGI FedCloud: Compute, Storage. AARC: AAI		
Infrastructure integration required			
Description of infrastructure integration activities (to be planned)	ACTIVITIES NEEDED: multiplication of server infrastructure (R servers, Geoservers,…) to increase performance; better integration of components by linking up components. Access regulated by central user authentication. Data flow from ship or sensor to servers; data stores, RDBMS, visualization interfaces, analytical workflows,… ALREADY IN PLACE (IN KIND): All components running on local servers; de-centralized authentication		
Training			
Description of training activities relevant to the service	Specific training activities would need to be developed for users		

2-Modelling Water Masses

Service overview	
Thematic Service name	**WATER MASSES MONITORING AND MODELLING**
Service description	This service aims to integrate different tools to manage the data that serves as input for Hydrodynamics and Water Quality Models, in particular those performed using Delft3D from Deltares. The use of those different tools will deploy a user-friendly and dynamic environment capable to exploit computing and storage resources over a cloud framework to do different actions: - Data Management. - Scenario and calibration performing for hydrodynamics and water quality modeling over a user friendly interface. - Models performing using distributed storage system. All the components involved in the service will integrate transversal tools to act coordinately. The service can be applied to different type of water mass, including rivers, lakes, reservoirs and oceans.
Service provider	LIFEWATCH ERIC SERVICE PROVIDED BY IFCA + ECOHYDROS SME
Service catalogue	LIFEWATCH MODELLING SERVICES
Value	- User-friendly access to computing resources. - Easy parameter sweep for calibrating models and scenario deployment in water masses (Hydrodynamics, Water Quality). - Access to computing and storage resources and dynamic staging of data (avoiding self-resources on laptop, PC...). - Integrated data and processing tools.
Current TLR level, acceptance criteria and validation/verification results	**TRL8** This service is set by different tools that are running and being used. The use of Delft3D modelling software is being done both in supercomputers and cloud computing resources within the context of different European projects and initiatives (ROEM+, INDIGO-DataCloud, Lifewatch).

	The monitoring data has been gathered since 2010 and the data visualization and management tool is running at IFCA (doriiie02.ifca.es). Presented at: - Delft3D Users Meeting (Delft, The Netherlands, 2015 and 2016). - INDIGO-DataCloud Review.
Access policy	*MARKET DRIVEN*
Terms of use	Available for LifeWatch users
User groups and scientific disciplines served	Biodiversity, Ecology, Ecosystem Research, Marine Research...
Service business model	Contracts with Water Authorities/ Third parties Ecohydros SL
Service architecture	

Service components

Name of component	Functional description, applicable standards and needed resource capacity (if applicable) e.g. CPU Time, storage capacity etc.	Provider If already appointed
EGI FEDCLOUD	Computing Resources Dynamic deployment of running instances (min. 4CPUs, 16GB RAM).	IFCA/IBERGRID
DELFT3D	Modelling Software	Deltares
MONITORING DATA	Data gathered from the water mass monitoring.	Ecohydros SL
OneData	Distributed Storage Service Min. 500GB disk per user. Rec. 1TB disk per user.	SW: INDIGO Space: IFCA/IBERGRID
Future GW	Friendly interface based on forms.	INDIGO
INDIGO	Service to dynamically deploy the	INDIGO

	Orchestrator	computing element.	
	Repositories	Repositories for Docker containers and job script to run the model.	Dockerhub Github
	INDIGO IAM	Authentication and Authorization service to access the components.	IAM
	Data Visualization and management Tool	Doriiie02.ifca.es	IFCA
Service integration with generic e-Infrastructures			
Integration activity and concerned service components	The Water Mass Modelling Service is being used both in Supercomputer and Cloud Computing resources. The monitoring is also using cloud resources. The following integration activities can create a unique environment: - Standard use of AAI solutions, working with all the different components. - Use of OneData as distributed storage solution. - Deployment of Data Management tools - Dynamic deployment of running components, using Virtual Machines or Dockers in FedCloud resources.		
List of requested service components	EGI FedCloud: Compute, Storage. AARC: AAI INDIGO: User Interface (Future Gateway), Software…		
Infrastructure integration			
Description of infrastructure integration activities (to be planned)	ACTIVITIES NEEDED: Monitoring platform ALREADY IN PLACE: HPC		
Training			
Description of training activities relevant (to be planned)	ACTIVITIES NEEDED: Specific Training ALREADY IN PLACE (IN KIND): *Tutorial EGU PICO, Environmental DataLab (Data Science Master)*		

3-Data Services

Service overview	
Thematic Service name	**Data Services**
Service description	1) One of the main characteristics of biodiversity data is its cross-disciplinary character and the extremely broad range of data types, structures, and semantic concepts which encompasses. **Data Services** is set of data services that: i) support ***cataloguing and publishing*** all the relevant meta-data information of the Greek biodiversity domain, ii) ***integrate data*** from heterogeneous sources by supporting the definitions of appropriate models, iii) efficiently ***discover biodiversity data*** of interest and enable the answering of complex queries that could not be answered from the individual sources. The **Data Services** allow the providers to express their metadata in a schema agnostic way; the provider is able to submit metadata according to their local format (e.g. Darwin Core) and these are automatically transformed with respect to the underlying centralized schemata of the infrastructure for gaining the advantages that semantic models offer. 2) The ***Data Services VRE*** (http://metacatalogue.portal.lifewatchgreece.eu/) which is accessible through the Lifewatch Greece Portal, provides access except of the afore mentioned services, to services of *data annotation, quality improvement, fundamental searching, semantic graph browsing, natural text descriptions production, and data manipulatio*n. It gives access to a Virtuoso Triple store which acts as the metadata repository and directory of the infrastructure, to an iRODS virtual file system which acts as the content storage, and to an annotation data base. For more information, you can read the relevant software description paper: Minadakis N, Marketakis Y, Doerr M, Bekiari C, Papadakos P, Gougousis A, Bailly N, Arvanitidis C (2016) LifeWatch Greece data-services: Discovering Biodiversity Data using Semantic Web Technologies. Biodiversity Data Journal 4: e8443. https://doi.org/10.3897/BDJ.4.e8443 or read the Data Services manual: http://metacatalogue.portal.lifewatchgreece.eu/getFile/documents/Data_Services_Tutorial.pdf

Service provider	**LifeWatch ERIC** Service provided by LifeWatchGreece from Institute of Marine Biology Biotechnology and Aquaculture (IMBBC) at Hellenic Centre for Marine Research (HCMR) with the cooperation of the **Foundation for Research and Technology Hellas (FORTH)**
Service catalogue	**Data Services** is part of LifeWatchGreece portal https://portal.lifewatchgreece.eu/
Value	The Data Services provide to the users / researchers an efficient way of discovering biodiversity data of interest coming from various sources and various formats by taking advantage of the semantic graphs capabilities. It enables complex querying execution and reasoning, answering to scientific questions that are difficult (or impossible) to be answered by the individual sources, saving a lot of time from the classical data discovery and access methods. Moreover, they support the publishing, annotation, cataloguing and integration of data, giving the ability to the researchers to make their data public, connected and widely distributed, taking the related credit and acknowledgment.
Current TRL level, acceptance criteria and validation/verification	TRL 7 in evolution to ***TRL8 to be completed in 2017*** (*system prototype demonstration in operational environment*). https://doi.org/10.3897/BDJ.4.e8443
Access policy	Using **Data Services** require users to register to LifewatchGreece portal.
Terms of use	For the time being, no terms of use are being enforced.
User groups and scientific disciplines	Academics, Researchers, Scientists, Students
Service business model	For the time being the Data Services is offered as a free service since its construction has been funded by taxpayers resources. However, its maintenance will inevitably need some additional resources. Thoughts have been expressed along the following lines: 1. A charge in those cases for which the use is targeted to industrial research and commits a great deal of its computational resources 2. Any additional development, under the scheme of joint venture.

<table>
<tr><td></td><td></td></tr>
<tr><td colspan="2">Service architecture</td></tr>
<tr><td>Service components</td><td>
<table>
<tr><th>Name of component</th><th>Functional description, applicable standards and resource capacity</th><th>Provider
If appointed</th></tr>
<tr><td>Directory</td><td>Virtuoso Triple Store, 128GB RAM, 1T storage capacity</td><td>HCMR</td></tr>
<tr><td>Metadata Repository</td><td>Virtuoso Triple Store,128GB RAM, 1T storage capacity</td><td>HCMR</td></tr>
<tr><td>Content Storage</td><td>iRODS, 128GB RAM, 4T storage capacity</td><td>FORTH</td></tr>
<tr><td>Fundamental Searching</td><td>Metaphacts Platform (JAVA, ReactJS), 128GB RAM, 1T storage capacity</td><td>Metaphacts, HCMR</td></tr>
<tr><td>Web Application</td><td>JSP, Javascript - Software</td><td>HCMR</td></tr>
<tr><td>Semantic Data Services</td><td>SOAP and REST services, JAVA API - Software</td><td>FORTH</td></tr>
</table>
</td></tr>
<tr><td colspan="2">Service integration with generic e-Infrastructures</td></tr>
<tr><td>Integration activity and concerned service components</td><td>- Transform every Semantic Service in REST service.
- Refine Data Service API.
- Migrate/integrate metadata repository, directory and content storage.
- Migrate/integrate fundamental search platform and annotation services.</td></tr>
<tr><td>List of requested service components</td><td>EGI FedCloud: Compute, Storage
AARC: AAI</td></tr>
<tr><td colspan="2">Infrastructure integration</td></tr>
<tr><td>Infrastructure integration activities (to be planned)</td><td>Data Services will use FedCloud computing and storage resources.</td></tr>
<tr><td colspan="2">Training</td></tr>
</table>

Description of training activities	Training workshops, Webinars, video-tutorials available.

4- GBIF data access under biogeographic context

Service overview	
Thematic Service name	GBIF data access under biogeographic context
Service description	This service aims at providing access with advanced facets to GBIF biodiversity data under a biogeographic context, using biogeographic regions as a domain for the data compilation, access, analysis and reporting, The services will be based on the Atlas of Living Australia platform, including modules for occurrences, species and species lists, regions, spatial portal (including spatial analysis), and will include convenient API services to support webapps.
Service provider	GBIF Portugal (Instituto Superior de Agronomia) GBIF Spain (Real Jardin Botánico) LIP IFCA
Service catalogue	Lifewatch Citizen Science Services
Value	Biodiversity data is freely available through GBIF, either for the global portal and services, or from national portals provided by the national nodes of GBIF. The latter access normally have the advantage of providing insights of the data to the user before the actual download, helping to decide on the relevance of the information for the purpose of the specific research. Nevertheless, subjects of research in biodiversity and ecology are not limited by political borders that normally define national portals, but are dependent of ecogeographical or biogeographical domains, which implies that the user does not have available a resource with an integrated view of the information at the scale of analysis, and has the burden of accessing, downloading and merging data from several sources. The implementation of a portal at the portal based on a biogeographical scale for the Iberian Peninsula, based on the Atlas of Living Australia, will allow the overcome those problems, providing an intuitive and ready to use service for assessment and use of biodiversity data

<table>
<tr><td>Current TLR level, acceptance criteria and validation/verification results</td><td colspan="3">TLR 8</td></tr>
<tr><td>Access policy</td><td colspan="3">Wide access: users can freely access to data and services provided, provided they agree with the GBIF data user agreement</td></tr>
<tr><td>Terms of use</td><td colspan="3">GBIF Data user agreement
http://www.gbif.org/terms/data-user</td></tr>
<tr><td>User groups and scientific disciplines served</td><td colspan="3">Biodiversity, biogeography, climate change, invasive species, food and farming, conservation, etc.</td></tr>
<tr><td colspan="4">Service architecture</td></tr>
<tr><td>Service components</td><td colspan="3"><table>
<tr><td>Name of component</td><td>Functional description, applicable standards and needed resource capacity (if applicable)
e.g. CPU Time, storage capacity etc.</td><td>Provider
If already appointed</td></tr>
<tr><td>Atlas of Living Australia platform</td><td>Virtual machines to support Apache Solr, Apache Cassandra, Apache Tomcat, Apache HTTP(S), Postgres, MySQL</td><td>Openstack cloud compute EGI FedCloud (IFCA)</td></tr>
</table></td></tr>
<tr><td colspan="4">Service integration with generic e-Infrastructures</td></tr>
<tr><td>List of requested service components</td><td colspan="3">API services available for data analysis by statistical package R provided by LifeWatch Competence Centre</td></tr>
<tr><td colspan="4">Infrastructure integration</td></tr>
<tr><td>Description of infrastructure integration activities relevant to the service (to be planned)</td><td colspan="3">Cloud compute support of Virtual Machines for the implementation of web, data and indexing services (e.g. Cassandra, Solr, Postgres, MySQL, Tomcat, Apache Web Server)</td></tr>
<tr><td colspan="4">Training</td></tr>
<tr><td>Description of training activities</td><td colspan="3">Training programs running by the Spanish GBIF node (http://www.gbif.es/formacion_in.php) and (soon) Portuguese</td></tr>
</table>

	GBIF Node

5- Citizen Science Services

Service overview	
Thematic Service name	**Citizen Science Services**
Service description	(1) Natusfera (http://natusfera.gbif.es/)A platform to publish, manage, and integrate biodiversity observations (species occurrences) (2) Automatic image analysis using AI and crowd-analysis (citizens' contributions). Taking as pilot the orchids scenario we can expand its scope with any set of scientific images (3) PyBossa: Crowdsourcing any task. Given a set of data (images, texts, audios) the service allows the creation of individual tasks that are delivered to visitors (4) Generating 3D models from pictures. Given a set of pictures of any natural resource (e.g. a tree, or a mountain) 3D models are automatically generated on the cloud Generating maps from pictures. Given a set of pictures of a given region (e.g. taken with drones for emergency management), maps are automatically generated on the cloud (5) CINDA (http://www.cinda.science/). An Open Science platform created for manage campaigns and contributions of volunteer networks.
Service provider	- *CREAF | Ecological and Forestry Applications Research Centre (Spain)* - *Spanish GBIF Node-CSIC* - *Instituto de Física de Cantabria (Spain)* - *BIFI-Unizar-Ibercivis(Spain)* - *iEcolab, IISTA, Univ. Granada, Univ. Córdoba*
Service catalogue	-Keep track: your natural observations and taxon lists, in the cloud -Identify species: Connect with other users to get help identifying what you see **-Gather information to identify the impacts of global change in order to design management actions that minimize them** -Create capacity, disseminate knowledges: Learn about nature through community support tools, increase your knowledge talking to other naturalists; Collaborate in interesting projects, or create your own -Support sciences: Help scientists find out where different taxa occur -Produce and make available different data products (maps, lists, summaries, analyses,..) relevant for the users
Value	*Enabling unprecedented sharing and re-use of information related to biological species and their geographic distributions for a variety of user communities: citizen scientists, land and protected area managers, scientists, etc.*

	Long-term monitoring based on the collaboration between citizens and scientists. *Increased number of potential users of the knowledge produced by the observatory.*
Current TLR level, acceptance criteria and validation/ verification results	*Provide information about* *(1) service status in terms of completeness and maturity (including link to relevant documentation),* **TRL 8-9: in production and mature** *(2) service acceptance criteria defined by customers and/or users (including e.g. aspects related to interoperability, availability, installability, performance, portability, recoverability, safety, scalability, usability;* Performance is satisfactory. Aspects such expanding the user base and enhancing interoperability needs further work. CINDA (Service 5) is multilingual, multiserver (it permits to connect to any server where CINDA has been installed) and multicampaign. *(3) results of validation and verification activities involving service providers and user communities.* For service 3, PyBossa automatically implements validation with work unit replication-quorum. Researchers can perform their own ad-hoc verification.
Access policy	Users are requested to register to publish and manage data and projects. Information tag as public by the respective contributors a freely available to anyone. Tools developed for these services are Open Source
Terms of use	For service 1: http://natusfera.gbif.es/pages/privacy (in Spanish) For service 5: http://www.cinda.science/open-data/ (in Spanish)
User groups and scientific disciplines served	Anyone with an interest in biodiversity from different avenues: ● Science ● Biodiversity conservation ● Land management ● Citizen science and civil initiatives ● Eco-tourism ● Recreation activities ● Etc.
Service business model	*Public funding, in the understanding that this service is a cost-effective system to gather manage and share biodiversity information relevant for science and essential for sound management and decision making regarding environmental issues (e.g. invasive species, protected species and areas, water quality, etc.)*

<table>
<tr><td colspan="2">Service architecture</td></tr>
<tr><td>Service components</td><td>
<table>
<tr><td>Name of component</td><td>Functional description, applicable standards and needed resource capacity</td><td>Provider
If already appointed</td></tr>
<tr><td>Identity and Access Management</td><td>AAI</td><td>IFCA</td></tr>
<tr><td>Cloud Compute</td><td>Computing Resources</td><td>IFCA</td></tr>
<tr><td>Online Storage</td><td>Storage Resources</td><td>IFCA</td></tr>
</table>
</td></tr>
<tr><td colspan="2">Service integration with generic e-Infrastructures</td></tr>
<tr><td>Integration activity and concerned service components</td><td>For service 1 (Natusfera):
-Under discussion: Building a gateway to allow data transfer and integration with the Global Biodiversity Information Facility (GBIF).
-Exploit images and identifications to build and refine computer-assisted species identification services based on machine learning methods</td></tr>
<tr><td>List of requested service components</td><td>Cloud Compute
Online Storage
Archive Storage
High-Throughput Compute</td></tr>
<tr><td colspan="2">Infrastructure integration</td></tr>
<tr><td>Description of infrastructure integration activities relevant to the service (to be planned)</td><td>Shot in the dark: developing software components to enable data transfer and integration with GBIF

Software development and ICT enhancing aimed to an image based computer-assisted species identification system</td></tr>
<tr><td colspan="2">Training</td></tr>
<tr><td>Description of training activities relevant</td><td>Training event aimed to citizen science organizations
Training event aimed to protected area managers
Workshop for parties interested in installing and deploying their own Natusfera platforms (using their own ICT infrastructure or the one already in place and dedicated to the current Natusfera site.</td></tr>
</table>

6- Image Classification Deep Learning Tools

Service overview	
Thematic Service name	IC-DLT (Image Classification Deep Learning Tool)
Service description	IC-DLT (Image Classification Deep Learning Tool) is a service designed to help users to train an image classificatory based on web. It is supported by different Neural Networks and Deep Learning tools and based on Python. The use of IC-DLT is a two-steps process: - First, the Neural Network is trained with the images datasets and tags provided by the user. This steps need access to GPUs. - Second, the trained Neural Network is connected with the web service that provides a web-based environment to upload new images to be classified. It returns a list of potential classification and accuracy (%).
Service provider	LIFEWATCH ERIC SERVICE PROVIDED BY IFCA
Service catalogue	LIFEWATCH SERVICES
Value	Easy system to train Neural Networks for image classification with no previous Deep Learning knowledge. Integrates software and scripts for perform in an easy way. Easy access to GPUs and Storage.
Current TLR level, acceptance criteria and validation/ verification results	IC-DLT is running in local resources of GPUs and it is being developed to work virtually over Docker containers to be easily capable to be deployed in different resource providers including EGI FedCloud. TLR Level: 7, evolution to TRL8 along 2017 (assured by contract in place) For more information regarding performance and use, check the GitHub Repository: https://github.com/IFCA/
Access policy	Free access for academic research under request
Terms of use	Check the manual
User groups and scientific disciplines	Students, Researchers, Scientists from: Biodiversity, Botanic, etc.
Service business model	*Free by now*

Service architecture

Service components

Name of component	Functional description, applicable standards and needed resource capacity (if applicable) e.g. CPU Time, storage capacity etc.	Provider If already appointed
EGI FEDCLOUD	Dynamic deployment of resources. - Model Training step: access to GPUs, Up to 1TB of storage per training. - Web-based classificatory: >4CPUs, >8GB memory, GPUs (recommendable but optional)	IFCA/ IBERGRID
Lasagne	http://lasagne.readthedocs.io	
Theano	http://deeplearning.net/software/theano/	
Python Libs	Present in Anaconda Distribution	
IC-DLT	Software	IFCA
INDIGO Orchestrator	Service to dynamically deploy the computing element.	INDIGO
Repositories	Repositories for Docker containers and job script to run the model.	Dockerhub Github
INDIGO IAM	Authentication and Authorization service to access the components.	IAM
Repositories	Repositories for Docker containers and job script to run the model.	Dockerhub Github
OneData	Distributed Storage Service 20 TB Disk	SW: INDIGO Space: IFCA/IBERGRID

Service integration with generic e-Infrastructures

Integration activity and concerned service components

ID-CLT is running on local resources and it has been tested with a training based on Data sets from PlantNet and validated with Portuguese Flora and iNaturalist images. The following integration activities can contribute to make it more scalable, easy to use and integrated:

- Standard use of AAI solutions, working with all the different components.
- Use of OneData as distributed storage solution.

	- Dynamic deployment of running components, using Virtual Machines or Dockers in FedCloud resources and exploitation of GPUs reources. - Use of Orchestrator to manage the 2-steps process: training and deployment of Image Classification web.
List of requested service components	EGI FedCloud: Compute (GPUs), Storage. AARC: AAI. INDIGO: Software (OneData, Orchestrator).
Infrastructure integration	
Description of infrastructure integration activities relevant service (to be planned)	IC-DLT will use FedCloud computing resources for running the training with GPUs and to deploy a web service for classification. A standardized AAI solution will be used (INDIGO IAM) and also a distributed storage system (INDIGO OneData). An Orchestration solution (INDIGO) is needed to manage the workflow and the dependencies between the step one (Training) and step two (Classification) that depends on the output of the first.
Training	
Description of training activities	A set of manuals and Video-tutorials will be available. Webinars and workshop will be organized as needed.

7-Genetic Services

Service overview	
Thematic Service name	**Genetic Services (Genetics vLab)**
Service description	1) The Genetics Services (or GeneticsvLab) is a new user friendly graphical interface for efficiency and ease in execution of QIIME (Quantitative Insights into Microbial Ecology). QIIME is an open - source bioinformatics pipeline for performing microbiome analysis from raw DNA sequencing data, designed to take users from raw sequencing data generated on the Illumina or other platforms through publication quality graphics and statistics. The Service focuses on taxonomic analysis from data derived from 454 Roche as well as Illumina sequencing technologies and provides the relevant software to perform analysis for both sequencing technologies 2) Genetics Services is running on a PC cluster (hosted by HCMR), using version 3.1.2 (2014-10-31) on a x86_64-pc-linux-gnu (64-bit) platform, and offers a virtual environment interface enabling users to perform analysis of microbial communities based on QIIME. 3) For more information, you can read the Genetics Services manual: https://qiime.portal.lifewatchgreece.eu/files/Genetics_services_manual.pdf
Service provider	LifeWatch ERIC Service provided by LifeWatchGreece from the Institute of Marine Biology, Biotechnology and Aquaculture (IMBBC) of the Hellenic Centre for Marine Research (HCMR)
Service catalogue	GeneticsvLab (https://qiime.portal.lifewatchgreece.eu)is part of the LifeWatchGreece portal (https://portal.lifewatchgreece.eu)
Value	The GeneticsvLab is designed to take users from raw sequencing data generated on the Illumina or other platforms through publication quality graphics and statistics.

<table>
<tr><td>Current TRL level, acceptance criteria and validation/verification results</td><td>TRL 7 – in evolution to TRL8 by end of 2017</td></tr>
<tr><td>Access policy</td><td>GeneticsvLab is available after registration at: https://portal.lifewatchgreece.eu/
A registered user can ask for access to GeneticsvLab for a specific period of time, which is granted almost immediately. This request is necessary to keep track for the number of active GeneticsvLab users.</td></tr>
<tr><td>Terms of use</td><td>A storage quota has been defined for each user's submitted jobs/analyses. Analyses results are stored in the user's workspace for a limited amount of time. These limitations are necessary in order to provide a minimum Quality of Service given the fact of limited storage resources. For details on the current policy please refer to:
https://www.lifewatchgreece.eu/?q=lifewatch-greece-documents</td></tr>
<tr><td>User groups and scientific disciplines served</td><td>Academics, Researchers, Scientists, Students</td></tr>
<tr><td>Service business model</td><td>For the time being the GeneticsvLab is offered as a free service since its construction has been funded by tax payers resources. However, its maintenance may need some additional resources. Thoughts have been expressed along the following lines:
1. A minimal charge for the purchase of the app for tablets and cell phones
2. Any additional development, under the scheme of joint venture.</td></tr>
<tr><td colspan="2">Service architecture</td></tr>
<tr><td>Service components</td><td>
<table>
<tr><th>Name of component</th><th>Functional description, applicable standards and needed resource capacity (if applicable) e.g. CPU Time, storage capacity etc.</th><th>Provider If already appointed</th></tr>
<tr><td>SAN</td><td>2 TB</td><td>HCMR</td></tr>
<tr><td>LifeWatchGreece Portal</td><td>Authentication Service</td><td>HCMR</td></tr>
</table>
</td></tr>
</table>

	HPC Cluster	200 cores, 1.5 TB RAM	HCMR
Service integration with generic e-Infrastructures			
Integration activity and concerned service components	Given sufficient resources, GeneticsvLab will integrate the full capabilities of QIIME pipeline and it will develop to incorporate other commonly used molecular biology algorithms.		
List of requested service components	Storage and Computational power will be requested in order to allow more users to work simultaneously and to achieve the overall improvement of GeneticsvLab.		
Infrastructure integration			
Description of infrastructure integration activities relevant to the proposed thematic service (to be planned)	GeneticsvLab services will use FedCloud computing and storage resources. Several transversal resources will be used based on different software providers (INDIGO, AARC).		
Training			
Description of training activities relevant to the proposed thematic service (to be planned in the project)	Training workshops, Webinars, video-tutorials available at Lifewatch ERIC web page.		

8- MiroCT vlab

Service overview	
Thematic Service name	**MicroCTvlab**
Service description	1) Microtomography (micro-computed tomography or micro-CT) is a method of non-destructive 3D x-ray microscopy, which allows the users to create 3D models of objects from a series of x-ray projection images, similar to the conventional clinical computer tomography. The Micro-CT virtual laboratory (Micro-CTvlab) offers to the user a collection of virtual image galleries of taxa, which can be displayed and disseminated through a web-based framework. With a few clicks, accurate, detailed and three-dimensional models of species can be studied and virtually dissected without destroying the actual specimen. The data and functions of the Micro-CT can be accessed either from a normal computer or through a dedicated version for mobile devices. 2) Micro-CT_{vlab} enables access to a collection of virtual 3D specimens which are annotated with metadata and can be interactively displayed and retrieved through a web-based application. 3) Keklikoglou K, Faulwetter S, Chatzinikolaou E, Michalakis N, Filiopoulou I, Minadakis N, Panteri E, Perantinos G, Gougousis A, Arvanitidis C (2016) Micro-CT: A web based virtual gallery of biological specimens using X-ray microtomography (micro-CT). Biodiversity Data Journal 4: e8740. https://doi.org/10.3897/BDJ.4.e8740 (http://bdj.pensoft.net/articles.php?id=8740)
Service provider	LifeWatch ERIC Service provided by LifeWatchGreece from Institute of Marine Biology Biotechnology and Aquaculture - IMBBC at Hellenic Centre for Marine Research – HCMR
Service catalogue	MicroCTvlab (http://microct.portal.lifewatchgreece.eu/) is part of LifeWatchGreece portal (https://portal.lifewatchgreece.eu)
Value	Experts in micro-CT technology can use the information in the MicroCTvlab to compare or discover protocols and scanning parameters for different species. Furthermore, members of the scientific community who are not yet familiar with this technology

	but work in areas such as taxonomy, evolutionary, developmental or functional biology could be attracted by the MicroCTvlab since this virtual service presents, through a range of examples, the potential for micro-CT imaging in many research fields. Natural history museums will naturally be highly interested in the MicroCTvlab and the underlying technology, since there is a need for massive digitisation and dissemination of natural history collections and this virtual lab could be used as a tool to achieve this. Furthermore, the MicroCTvlab can be used for educational purposes since it offers information on the morphology and anatomy of species and the 3D model scan be interactively manipulated by the students.
Current TLR level, acceptance criteria and validation/verification results	**TRL 8 (system complete and qualified)** *A standardised workflow has been developed for the creation of micro-CT datasets, protocols and terms for documenting each dataset with metadata[1], and a web based environment for the publication, dissemination and on-the-fly rendering manipulation of these datasets and their metadata (https://microct.portal.lifewatchgreece.eu/).* *[1] Faulwetter, S., Minadakis, N., Keklikoglou, K., Doerr, M., & Arvanitidis, C. First steps towards the development of an integrated metadata management system for biodiversity-related micro-CT datasets.*
Access policy	For the time being, no terms of use are being enforced.
Terms of use	The source code for the mobile application is licensed under MIT license. For the web application the source code is licensed under the GNU General Public License. The content of the MicroCT is available under a Creative Commons Attribution License (CC-BY) unless indicated otherwise
User groups and scientific disciplines	Scientists, Researchers, Students, Artists, Animators
Service business model	For the time being the MicroCTvlab is offered as a free service since its construction has been funded by taxpayers' resources. However, its maintenance will inevitably need some additional resources. Thoughts have been expressed along the following lines: 1. A minimal charge for the purchase of the app for tablets and cell phones 2. A charge in those cases for which the use is targeted to industrial research and commits a great deal of its

<table>
<tr><td></td><td>computational resources
3. Any additional development, under the scheme of joint venture.</td></tr>
<tr><td colspan="2">Service architecture</td></tr>
<tr><td>Service components</td><td>
<table>
<tr><td>Name of component</td><td>Functional description, applicable standards and needed resource capacity (if applicable) e.g. CPU Time, storage capacity etc.</td><td>Provider
If already appointed</td></tr>
<tr><td>Data Storage</td><td>50TB per year</td><td></td></tr>
</table>
</td></tr>
<tr><td colspan="2">Service integration with generic e-Infrastructures</td></tr>
<tr><td>Integration activity and concerned service components</td><td>The service integration activities are related to: a) the large size of the Micro-CT datasets constitutes a restriction for native integration and upload of the raw data (i.e. the high-resolution cross-section datasets). Currently, several datasets are available only through external links to the Dryad data repository, but an installation of a storage area network (SAN) is planned to overcome this restriction. With this SAN in place, all datasets will be made available for download; b) a service needs to be developed to
allow other Micro-CT users to submit and share their raw data through the Micro-CT; c) the communication with the LifeWatchGreece data services catalogue needs to be improved to allow refined querying for datasets; d) the process of creating preview files, descriptions and .nifti files for online manipulation needs to be automated so that the integration of additional datasets can be achieved more quickly.</td></tr>
<tr><td>List of requested service components</td><td>EGI FedCloud: Storage capacity will be requested as the size of the micro-CT datasets range from 1 to 50GB each</td></tr>
<tr><td colspan="2">Infrastructure integration</td></tr>
<tr><td>Description of infrastructure integration activities</td><td>MicroCTvlab could use FedCloud storage resources.</td></tr>
<tr><td colspan="2">Training</td></tr>
<tr><td>Description of</td><td>Training workshops related to digitization methods, Webinars,</td></tr>
</table>

training activities	video-tutorials available at Lifewatch ERIC web page.

9- R Services

Service overview	
Thematic Service name	**RvLab & R services**
Service description	1) The RvLab in its current version makes use of the VEGAN library of the "R" package, which is a freely available language and environment for statistical computing, widely used by scientists working in many research disciplines. It supports an integrated and optimized online R environment. This vLab tackles common problems faced by R users, such as limited computational capacity when running the package on a single core. Many of the routines operating in the R environment, such as certain multivariate analyses, are often of high computational demand and cannot deliver a result when the respective datasets are in the form of large matrices or the algorithms engage iterative routines for calculation. This vLab allows for a predefined, commonly used set of R functions to run on the LifeWatchGreece Research Infrastructure in order to support large-scale computational and modeling activities. The vLab is also available as a mobile/tablet application. 2) RvLab is running on a PC cluster (hosted by HCMR), using version 3.1.2 (2014-10-31) on a x86_64-pc-linux-gnu (64-bit) platform, and offers an intuitive virtual environment interface enabling users to perform analysis of ecological and microbial communities based on optimized vegan functions. The RvLab will integrate other R-based services from Lifewatch partners. In particular, Lifewatch Belgium (VLIZ) is offering Rshiny applications for data analysis and Lifewatch Spain (IFCA) provides Jupyter portal that runs both R and python scripts and programs for analysis. 3) For more information, you can read the relevant software description paper: Varsos C, Patkos T, Oulas A, Pavloudi C, Gougousis A, Ijaz U, Filiopoulou I, Pattakos N, Vanden Berghe E, Fernández-Guerra A, Faulwetter S, Chatzinikolaou E, Pafilis E, Bekiari C, Doerr M, Arvanitidis C (2016) Optimized R functions for analysis of ecological community data using the R virtual laboratory (RvLab). Biodiversity Data Journal 4: e8357. https://doi.org/10.3897/BDJ.4.e8357 or watch the RvLab's video tutorial: https://youtu.be/87nw-8W6myI or read the RvLab's manual:

	https://rvlab.portal.lifewatchgreece.eu/files/RvLab_manual.pdf
Service provider	LifeWatch ERIC RvLab is provided by LifeWatchGreece (HCMR). Rshiny applications provided by Lifewatch Belgium (VLIZ). Jupyter environment provided by Lifewatch Spain (IFCA).
Service catalogue	RvLab (https://rvlab.portal.lifewatchgreece.eu/) is part of the LifeWatchGreece portal (https://portal.lifewatchgreece.eu/) and the LifeWatch Marine VRE (http://marine.lifewatch.eu/lifewatch-greece-rvlab) RvLab, Rshiny applications and Jupyter environment are part from LifeWatch Analytical services and Virtual Research Environments.
Value	RvLab is a very useful and powerful tool, both for users who are already familiar with R (and some of its functions) but also for students and/or scientists who are in favour of open source software and would like to dedicate some time to get familiar with its functions, without having to go through the steep command line R learning curve. Performing analysis with high computational or memory requirements that would not be feasible for a personal computer, is now possible for any subscribed user. The results are made available asynchronously. In addition, the user can run several functions at the same time, without having to wait for the first analysis (R function) to finish in order to submit a second analysis. The user can benefit from the availability of newly designed functions if the dataset to be analysed requires their implementation. In addition, the RvLab is an interactive virtual laboratory; should the user require other types of functions, these can be added in the "laboratory" and become available online in a short time. All LifeWatch R-services provide a user-friendly environment that allows the users to access powerful computing resources.
Current TRL level, acceptance criteria and validation/ verification results	RvLab is running on a PC cluster (hosted by HCMR), using version 3.1.2 (2014-10-31) on a x86_64-pc-linux-gnu (64-bit) platform since Rshiny is currently running at VLIZ resources and is being used to exploit Marine Observatory data. Jupyter is deployed at IFCA and connected to Supercomputer resources. **TRL 8 – systems complete and qualified**

<table>
<tr><td></td><td></td></tr>
<tr><td>Access policy</td><td>Wide-access: RvLab is available, after registration, at: https://portal.lifewatchgreece.eu/ A registered user can ask for access to RvLab for a specific period of time, which is granted almost immediately. This request is necessary to keep track for the number of active RvLab users.
Rshiny and Jupyter services are available for researchers by request.</td></tr>
<tr><td>Terms of use</td><td>For the time being, no terms of use are being enforced.</td></tr>
<tr><td>User groups and scientific disciplines served</td><td>Researchers, Scientists, Students, Environmental managers, Environmental Stakeholders
Biodiversity, Marine Research, Ecology</td></tr>
<tr><td>Service business model</td><td>For the time being the RvLab is offered as a free service since its construction has been funded by taxpayers’ resources. However, its maintenance will inevitably need some additional resources. Thoughts have been expressed along the following lines:
1. A minimal charge for the purchase of the app for tablets and cell phones
2. A charge in those cases for which the use is targeted to industrial research and commits a great deal of its computational resources
3. Any additional development, under the scheme of joint venture.</td></tr>
<tr><td colspan="2">Service architecture</td></tr>
<tr><td>Service components</td><td>
<table>
<tr><th>Name of component</th><th>Functional description, applicable standards and needed resource capacity (if applicable) e.g. CPU Time, storage capacity etc.</th><th>Provider If already appointed</th></tr>
<tr><td>SAN</td><td>2 TB</td><td>HCMR</td></tr>
<tr><td>EGI FedCloud</td><td>200 cores, 1.5 TB RAM</td><td>HCMR</td></tr>
<tr><td>LifeWatchGreece Portal</td><td>Authentication Service</td><td>HCMR</td></tr>
<tr><td>INDIGO AAI</td><td>Authentication Service</td><td>INDIGO</td></tr>
<tr><td>Data Storage</td><td>20TB + 5TB per year (Marine</td><td></td></tr>
</table>
</td></tr>
</table>

		Observatory)	
	R libraries		
	R Shiny	Software	RStudio Inc.
	Jupyter	Software	Project Jupyter
Service integration with generic e-Infrastructures			
Integration activity and concerned service components	- Transform R vLab to an API-based service. - Deployment of a single-point access for R services. - Integration of Data Storage for R vLab. - Integration to HPC cluster for R vLab. - Integration to Cloud AAI based on Open standards like OpenID connect. - Integration of other R services: Rshiny, Rstudio, Jupyter.		
List of requested service components	EGI FedCloud: Compute, Storage AARC: AAI		
Infrastructure integration			
Description of infrastructure integration activities relevant to the proposed thematic service (to be planned)	RvLab and R Lifewatch services will use FedCloud computing and storage resources. Several transversal resources will be used based on different software providers (INDIGO, AARC).		
Training			
Description of training activities relevant to the service.	Training workshops, Webinars, video-tutorials available at Lifewatch ERIC web page.		

10- Semantic Tools

Service overview	
Thematic Service name	*Semantic tools for aquatic functional ecology*
Service description	This service provides different thesauri on functional traits of aquatic organisms (phytoplankton, zooplankton, macrozoobenthos, macroalgae and fish). The use of thesauri is an acknowledged good practice to establish the foundation for semantic interoperability, a critical requirement for reuse and sharing of data. Thesauri, in fact, collectively constructed, bypass ambiguity issues in natural language, facilitating the identification and integration of the information available in multiple data sources and allowing both scientists and computer applications to interpreter more effectively the meaning of data. Semantic technologies provide a promising way to properly describe and interrelate different data sources in ways that reduce barriers to data discovery, integration, and exchange among ecological resources and researchers. LifeWatch Functional Traits Thesauri focus specifically to harmonization and integration of individual body-size data that serve as input for collaborative experiments on the Metabolic Scaling Theory.
Service provider	*LIFEWATCH ERIC* *SERVICE PROVIDED BY LifeWatch-ITA*
Service catalogue	LIFEWATCH Catalogue of Services (http://www.servicecentrelifewatch.eu/catalogue-of-services)
Value	*It will reduce barriers to data discovery, integration, and exchange among ecological resources and researchers*
Current TRL level	**TLR9**
Access policy	Wide access: users can freely access the service provided;

<table>
<tr><td>Terms of use</td><td>For the time being, no terms of use are being enforced</td></tr>
<tr><td>User groups and scientific disciplines</td><td>Biodiversity and Ecology</td></tr>
<tr><td>Service business model</td><td>For the time being this thematic service is offered as a free service.</td></tr>
<tr><td colspan="2">Service architecture</td></tr>
<tr><td>Service components</td><td>
<table>
<tr><th>Name of component</th><th>Functional description, applicable standards and needed resource capacity</th><th>Provider
If already appointed</th></tr>
<tr><td>MYSQL</td><td>DBMS</td><td>LW-ITA</td></tr>
<tr><td>ThemaTres</td><td>Thesauri Management Software and SPARQL EndPoint</td><td>LW-ITA</td></tr>
<tr><td>OpenStack</td><td>OpenSource Cloud Management Framework, enhanced with INDIGO and EGI addition. (Providing about 50 CPU/Core, 100GB or RAM and 20TB of disc space)</td><td>Public Available + INDIGO / EGI (LW ITA)</td></tr>
<tr><td>INDIGO-IAM</td><td>The INDIGO IAM service provides a layer where identities, enrolment, group membership, attributes and policies to access distributed resources and services can be managed in a homogeneous and interoperable way. It supports the federated authentication mechanisms behind the INDIGO AAI.
The IAM service provides user identity and policy information to services so that consistent authorization decisions can be enforced across distributed services.</td><td>INDIGO (LW ITA)</td></tr>
</table>
</td></tr>
</table>

		Identity and Access Management is provided through multiple methods (SAML, OpenID Connect and X.509) by leveraging on the credentials provided by the existing Identity Federations (i.e. IDEM, EDUGAIN).	
	INDIGO PaaS Layer	Provide the federation across several IaaS together with the capability of data-aware scheduling. The INDIGO PaaS is also able to provide the needed automation that could implement also complex cluster of services. The PaaS µServices are also able to implement auto-provisioning of resource based on the usage of the services. (Using about 20 CPU/Core, 50GB or RAM and 4TB of disc space on a OpenStack instance)	INDIGO (LW ITA)
Service integration with generic e-Infrastructures			
Integration activity and concerned services	Integrating the high-level services described in the INDIGO PaaS, in order to obtain automatic scalability and high availability.		
List of requested service components	EGI FedCloud: Compute, Storage. INDIGO: AAI, INDIGO-PaaS		
Infrastructure integration			
Description of infrastructure integration activities	ACTIVITIES NEEDED: Integration of the high-level services in the INDIGO PaaS. ALREADY IN PLACE (IN KIND): Cloud infrastructure, INDIGO IAM		
Training			

Description of training activities	IN PLACE: *Tutorial available on the LifeWatch e-Training Platform*

11- Phytoplankton VRE

Service overview	
Thematic Service name	**Phytoplankton Virtual Research Environment**
Service description	This service offers: basic, fundamental information on morphological and taxonomic characterization of phytoplankton through a taxonomic Atlas of transitional and coastal water phytoplankton and Atlas of phytoplankton shape, which are independent and interoperable; a data template format for data standardization and harmonization; machine to machine tools to estimate quantitatively morphological and demographic traits and cell size-based multi-metric descriptors of the ecological status of aquatic ecosystems; data analysis and modeling. Going in the management data tool, users can access an interface that allows them to perform a number of queries on the Phytoplankton database and obtain an exhaustive data export according to designed aims The framework may be operated by phytoplanktonologist with various level of knowledge, from basic to advanced levels of knowledge. The service provides both a procedural method and web-based working environment to assist and help a phytoplanktonologist step by step, from taxonomic identification, specific shape association, measurement of linear dimensions characterizing each single shape, to morphological and demographic traits computation. A user can find easy and transparent access to both a set of computational tools and sharing data facility. The harmonization and standardization of data using a data template ensure the consistency and the usefulness of phytoplankton database, facilitating their comparability and accessibility and enhancing the information exchange between agencies.
Service provider	LIFEWATCH ERIC SERVICE PROVIDED BY LIFEWATCH ITA
Service catalogue	LIFEWATCH CATALOGUE OF SERVICES
Value	Very useful and fast guide helping in the identification, shape association and detailed computational processes of phytoplankton studies. Easy use of several tools in a more reliable way to analyze, compare and share the data obtained by comparable methodology, following a data template harmonization.
Current TLR level,	(1) TLR9

<table>
<tr><td>acceptance criteria and validation/verification results</td><td>(2) the services are interoperable with the LW DATA & METADATA STANDARD MODEL
(3) REFERENCES</td></tr>
<tr><td>Access policy</td><td>Policy based</td></tr>
<tr><td>Terms of use</td><td>Common citation format like BBTExt, MLA or APA, DOI, URL</td></tr>
<tr><td>User groups and scientific disciplines served</td><td>Biodiversity, Ecology, Ecosystem Research, Marine and Transitional systems</td></tr>
<tr><td colspan="2">Service architecture</td></tr>
<tr><td>Service components</td><td>
<table>
<tr><th>Name of component</th><th>Functional description, applicable standards and needed resource capacity</th><th>Provider
If already appointed</th></tr>
<tr><td>Data store: MS SQL SERVER</td><td>Data from monitoring programs, research projects</td><td>LW ITA</td></tr>
<tr><td>Data store: MONGODB</td><td>Storage Service</td><td>LW ITA</td></tr>
<tr><td>Data PORTAL & Service Catalogues</td><td>Data from Lifewatch Italy JRU, based on Liferay Portlet container framework</td><td>LW ITA</td></tr>
<tr><td>R Server</td><td>Computational Service</td><td>LW ITA</td></tr>
<tr><td>R Shiny</td><td>Friendly interface based on forms.</td><td>LW ITA</td></tr>
<tr><td>MongoDB</td><td>Storage Service</td><td>LW ITA</td></tr>
<tr><td>VRE</td><td>HYPER-V Virtualization Framework</td><td>LW ITA</td></tr>
<tr><td>Data PORTAL & Service Catalogues</td><td>Data from Lifewatch Italy JRU, based on Liferay Portlet container framework</td><td>LW ITA</td></tr>
<tr><td>OpenStack</td><td>OpenSource Cloud Management Framework, enhanced with INDIGO and EGI addition.
(Providing about 50 CPU/Core, 100GB or RAM and 20TB of disc</td><td>Public Available + INDIGO / EGI (LW ITA)</td></tr>
</table>
</td></tr>
</table>

	space)	
INDIGO-IAM	The INDIGO IAM service provides a layer where identities, enrolment, group membership, attributes and policies to access distributed resources and services can be managed in a homogeneous and interoperable way. It supports the federated authentication mechanisms behind the INDIGO AAI. The IAM service provides user identity and policy information to services so that consistent authorization decisions can be enforced across distributed services. Identity and Access Management is provided through multiple methods (SAML, OpenID Connect and X.509) by leveraging on the credentials provided by the existing Identity Federations (i.e. IDEM, EDUGAIN, etc.).	INDIGO (LW ITA)
INDIGO PaaS Layer	Provide the federation across several IaaS together with the capability of data-aware scheduling. The INDIGO PaaS is also able to provide the needed automation that could implement also complex cluster of services. The PaaS µServices are also able to implement auto-provisioning of resource based on the usage of the services. (Using about 20 CPU/Core, 50GB or RAM and 4TB of disc space on a OpenStack instance)	INDIGO (LW ITA)

Service integration with generic e-Infrastructures

Integration activity and concerned service components	Integrating the high-level services described in the INDIGO PaaS, in order to obtain automatic scalability and high availability.
List of requested service components	EGI FedCloud: Compute, Storage. INDIGO: AAI, INDIGO-PaaS
Infrastructure integration	
Description of infrastructure integration activities (to be planned)	ACTIVITIES NEEDED: Integration of the high-level services in the INDIGO PaaS. ALREADY IN PLACE (IN KIND): Cloud infrastructure, INDIGO IAM
Training	
Description of training activities	ALREADY IN PLACE (IN KIND): Tutorial available on the LifeWatch e-Training Platform

12- Ecological Data Analysis Platform

Service overview	
Thematic Service name	**Platform for Ecological Data Analysis**
Service description	This service aims to share, harmonize, integrate and reuse observatory data, offering different tools for data management and statistical tools to support the analysis of ecological data. Metadata and data are managed by user-friendly interface and tools to describe and detail information about data and to check its quality (format, numeric and taxonomic). The produced data can directly be imported in a Virtual Research Environment which provide workflow management systems such as Taverna and Galaxy, or G.I.S. systems and statistical tools (i.e. R, Matlab, etc.) to analyse data. More over the Virtual Research Environments (VREs) enable dynamic collaboration promoting community engagement. The service can be applied to different type of marine, freshwater and terrestrial ecological communities.
Service provider	*LIFEWATCH ERIC* *SERVICE PROVIDED BY LIFEWATCH – ITA*
Service catalogue	*LIFEWATCH SERVICE CENTRE CATALOGUE OF SERVICES*
Value	User friendly, easy parameter sweep, possibility to upload the own data users, access to LifeWatch computing resources and dynamic staging of data (avoiding self-resources on laptop, PC...). Integrated data and processing tools.
Current TRL level	(1) TRL9 (2) the services are interoperable with the LW DATA & METADATA STANDARD MODEL
Access policy	POLICY-BASED
Terms of use	Common citation format like BBTExt, MLA or APA, DOI, URL
User groups and scientific disciplines served	Biodiversity, Ecology, Ecosystem Research, Monitoring...

Service architecture

Service components

Name of component	Functional description, applicable standards and needed resource capacity	Provider If already appointed
Biodiversity partitioning	Biodiversity partitioning across nested spatial level. Given as input a dataset of biotic measurements (species occurrences) and nested spatial levels (e.g. Sites within Locations within Regions), it allows to compare the biodiversity partitioning across higher and lower levels.	LW ITA
ISS_benthos	Calculation of macrozoobenthic ISS index (Basset et al. 2012, Barbone et al. 2012). Given as input a dataset with macrozoobenthic individual observations (individuals occurence and body size), produce as output a list of ISS values per site.	LW ITA
ISS_phyto	Calculation of phytobenthos ISS index (Vadrucci et al. 2012). Given as input a dataset with phytoplankton individual observations (individual occurrence and body size), produce as output a list of ISS values per site	LW ITA
Niche_filtering	Suite of logistic and quantile regression models for the analysis of taxa responses (both in terms of occurrences and abundances) to environmental condition(s). Given as input a dataset comprehensive of biotic (taxa abundances or occurrences) and abiotic (environmental parameters), it returns as output parametrized niche models (both as tables and figures). It must be noted that logistic regression is based on pseudo-absences, that can be assessed only on multi-taxa dataset.	LW ITA

Taxonomic rarefaction	Evaluation of the estimated number of taxa for given sample size. Given as input a dataset of biotic measurements (species occurrences), it allows to evaluate the relation between sampling size and with rarefaction curves (both table and graph) and simulated species richness	LW ITA
ActionMed Indicators Catalogue	User-friendly, electronic catalogue of environmental indicators applicable to assess ecosystem health in marine waters	LW ITA
Data PORTAL & Service Catalogues	Data from LifeWatch Italy JRU, based on Liferay Portlet container framework	LW ITA
R Server	Computational Service	LW ITA
R Shiny	Friendly interface based on forms.	LW ITA
MongoDB	Storage Service	LW ITA
VRE	HYPER-V Virtualization Framework	LW ITA
OpenStack	Open Source Cloud Management Framework, enhanced with INDIGO and EGI addition. (Providing about 50 CPU/Core, 100GB or RAM and 20TB of disc space)	Public Available + INDIGO / EGI (LW ITA)
INDIGO-IAM	The INDIGO IAM service provides a layer where identities, enrolment, group membership, attributes and policies to access distributed re sources and services can be managed in a homogeneous and interoperable way. It supports the federated authentication mechanisms behind the INDIGO AAI. The IAM service provides user identity and policy information to services so that consistent authorization decisions can be enforced across distributed services	INDIGO (LW ITA)

		Identity and Access Management is provided through multiple methods (SAML, OpenID Connect and X.509) by leveraging on the credentials provided by the existing Identity Federations (i.e. IDEM, EDUGAIN, etc).	
	INDIGO PaaS Layer	Provide the federation across several IaaS together with the capability of data-aware scheduling. The INDIGO PaaS is also able to provide the needed automation that could implement also complex cluster of services. The PaaS µServices are also able to implement auto-provisioning of resource based on the usage of the services. (Using about 20 CPU/Core, 50GB or RAM and 4TB of disc space on a OpenStack instance)	INDIGO (LW ITA)
Service integration with generic e-Infrastructures			
Integration activity and concerned service components	Integrating the high-level services described in the INDIGO PaaS, in order to obtain automatic scalability and high availability.		
List of requested service components	EGI FedCloud: Compute, Storage. INDIGO: AAI, INDIGO-PaaS		
Infrastructure integration			
Description of infrastructure integration activities	ACTIVITIES NEEDED: Integration of the high-level services in the INDIGO PaaS. ALREADY IN PLACE (IN KIND): Cloud infrastructure, INDIGO IAM,		
Training			
Description of training activities	ALREADY IN PLACE (IN KIND): *Tutorial available on the LifeWatch e-Training Platform*		

13- Digital Knowledge Preservation Framework

Service overview	
Thematic Service name	Lifewatch Digital Knowledge Preservation Framework
Service description	The Digital Knowledge Preservation Platform is a tool for Open Data supporting the full research data life cycle. It is currently being developed at IFCA (Instituto de Física de Cantabria) as a combination of different extended tools: DMPTool (https://dmptool.org/) with pilot semantics features (RDF export, parameters definition), INVENIO customized version to fulfill the entire research data life cycle and Jupyter (http://jupyter.org/) as processing tool and reproducible environment. This complete platform aims to provide an integrated environment for research data management warranting the FAIR+R principles: • Findable: The Web portal provides a search engine and all elements including metadata to make them easily findable. • Accessible: The elements are available online with both internal PIDs and DOIs provided by Datacite. • Interoperable: Datasets can be combined to perform new analysis. OAI-PMH standard will be implemented. • Re-usable: different licenses types and embargo periods available to be defined. • +Reproducible: Integrated to cloud computing resources (Jupyter and Virtual Machines). The deployment of the entire system over a Cloud framework helps to build a dynamic and scalable solution, not only for storing data but also as a useful tool for the final user, who is able to process and analyze the data.
Service provider	LIFEWATCH ERIC SERVICE PROVIDED BY IFCA
Service catalogue	LIFEWATCH DATA SERVICES
Value	Set of integrated software and tools to manage the whole data life cycle. Tracking and registry of all the connected stages, warranting reproducibility of the experiment.
Current TLR level,	An instance of the last released version of the portal is running at IFCA

acceptance criteria and validation/verification results	with FedCloud resources since 2015. The new version is on development, with new features being added. TLR Level: 8 https://github.com/IFCA/lifewatch_osf Link: https://193.146.75.147
Access policy	Free access for LifeWatch members
Terms of use	Check the manual
User groups and scientific disciplines served	The current version is LifeWatch-oriented, but can be expanded to Students, Researchers, Scientists from: Life Sciences, Biodiversity, any other field with Data Management requirements.
Service business model	Define estimated cost of ownership/year, funding streams and sustainability plans.
Service architecture	

Service components

Name of component	Functional description, applicable standards and needed resource capacity (if applicable) e.g. CPU Time, storage capacity etc.	Provider If already appointed
EGI FEDCLOUD	Dynamic deployment of resources. - Portal: At least one instance with (8 CPUs, 8GB). 20 TB disk. Recommended: replica. - Analysis instances: Up to 256 cores. Up to 32 GB memory. On demand.	IFCA/IBERGRID
Invenio	http://invenio-software.org	Invenio
DMP Tool	https://**dmptool**.org	University of California
Jupyter	http://jupyter.org	Project Jupyter
INDIGO IAM	Authentication and Authorization service to access the components.	IAM
Repositories	Repositories for Docker containers and code	Dockerhub Github
OneData	Distributed Storage Service 20 TB Disk	SW: INDIGO Space: IFCA/IBERGRID

	Lifewatch Digital Knowledge Preservation Framework	https://github.com/IFCA/lifewatch_osf	IFCA
Service integration with generic e-Infrastructures			
Integration activity and concerned service components	One instance of the last portal release is running at IFCA. The portal developer's team is currently developing the new version to add some new features. The new version needs a set of integration actions, not only for the portal components but also with some other external. - Standard use of AAI solutions, working with all the different components. - Use of OneData as distributed storage solution. - Dynamic deployment of running components for analysis, using Virtual Machines or Dockers in FedCloud resources. - Communication between APIs from components.		
List of requested service components	EGI FedCloud: Compute, Storage. AARC: AAI. INDIGO: Software (IAM, OneData).		
Infrastructure integration			
Description of infrastructure integration activities	The portal will use FedCloud computing resources for running the analysis step. A standardized AAI solution will be used (INDIGO IAM, AARC) and also a distributed storage system (INDIGO OneData).		
Training			
Description of training activities	A set of manuals and Video-tutorials will available in the documentation. Webinars and workshop will be organized on demand.		

14- Remote Monitoring and Smart Sensing

Service overview	
Thematic Service name	Remote Monitoring and Smart Sensing Analysis Service
Service description	The Remote Monitoring and Smart Sensing Analysis Service is a webserver designed to cover the entire process (from the selection, downloading to the view and analysis) required to work with Sentinel data products. First, the webserver provides an Interface to search, find and download Copernicus Sentinel satellite products easily, and after then provides different tools to manage and work with the products. During the downloading process, the user can perform a valid search for different zones, and also restrict the queries by different keywords: cloud coverage, date, platform name (S1, S2, S3). In case of interruptions or other exceptions, downloading will restart from where it left off. At the same time, a geospatial integration with Smart Sensing data (where applicable, mainly from isolated areas) will be performed In terms of data treatment, the following products are available to be processed: - Sentinel-1 data products. Using these data products, the user is allowed to open and extract some information and metadata. Users also can Pre-process (calibrate and perform geometric corrections) images to see the clean image, and also produce subsets and plot them using the Toolbox provided by Copernicus (SNAP). - Sentinel-2 data products. Using these data products, the user is allowed to open and get the following product information and metadata: size, available bands for analysis, etc. Plotting an image with a single band or a RGB image by composition bands. The user is also allowed to generate three vegetation indices. NDVI, NDI45, GNDVI. Since the original image is a set of tiles of 100 km^2, the user can work with one of those tiles or select a subset within the product defining a polygon. - Sentinel-3 data products. Although these data products are not yet available, they will allow the user to get data from water masses in order to generate information like Algae Bloom status.

Service provider	LIFEWATCH ERIC SERVICE PROVIDED BY IFCA
Service catalogue	LIFEWATCH MONITORING SERVICES
Value	Set of integrated software to get access to Copernicus data products and tools for analyze them. Access to computing and storage resources.
Current TLR level, acceptance criteria and validation/verification results	The current version of the Remote Monitoring and Smart Sening service for satellite images analysis is deployed in a testbed environment using FedCloud resources. The main functions are programmed in python and use Jupyter as IDE and running environment. TLR Level: 7 in evolution to TRL8 (to be completed in 2017, 2 FTE working on it) For more information regarding performance and use: Manual, Code: https://github.com/IFCA/
Access policy	Free access for academic researchers under request
Terms of use	Check the manual Copernicus Terms of use: https://sentinels.copernicus.eu/web/sentinel/terms-conditions
User groups and scientific disciplines served	Students, Researchers, Scientists from: Biodiversity, Environmental Research, Hydrology. Close link with the development of River Basins, Deltas and Harbors-Ports applications (particularly through existing connections with DANUBIUS-RI).
Service architecture	

Service components

Name of component	Functional description, applicable standards and needed resource capacity	Provider If already appointed
EGI FEDCLOUD	Minimum requirements: - 8 GB RAM - 8 CPUs ~ 3.0GHz - 500 GB Disk Recommended: - 16 GB RAM - >8 CPUs ~ 3.0GHz - 1TB SSD Disk	IFCA/IBERGRID

	Copernicus Sentinel Data Products	Data products Sentinel 1, 2, 3. https://sentinels.copernicus.eu	ESA
	Jupyter	http://jupyter.org/	Project Jupyter
	Analysis Software	Software for managing data products and processing. Python	IFCA
	INDIGO IAM	Authentication and Authorization service to access the components.	INDIGO
Service integration with generic e-Infrastructures			
Integration activity and concerned service components	The Remote Monitoring Analysis Service is running at IFCA with local users. The integration activities will allow other users from LifeWatch and other user communities to use the service and exploit EGI FedCloud Computing and storage resources. - Standard use of AAI solutions, working with all the different components. - Exploitation of EGI FedCloud resources. - DevOps oriented deployment of the system. - Monitoring and Resource management.		
List of requested service components	EGI FedCloud: Compute, Storage. AARC: AAI. INDIGO: IAM		
Infrastructure integration			
Description of infrastructure integration activities	The Remote Monitoring Analysis Service will use FedCloud computing resources for instance deployment and for running the different images analysis. A standardized AAI solution will help to deploy a more standardized system (INDIGO IAM or AARC solutions).		
Training			
Description of training activities	Webinars or workshops will be organized as needed.		

15- TRUFA

Service overview	
Thematic Service name	TRUFA (Transcriptomes User-Friendly Analysis)
Service description	TRUFA (Transcriptomes User-Friendly Analysis) is a free webserver designed to help you perform RNA-seq analysis. So far, TRUFA is allowing you to execute the following steps (programs used are specified in the parentheses): Reads cleaning: • Reads quality control (FASTQC) • Quality trimming and duplicates removal (Prinseq) • Trimming adapters (Cutadapt) • Filtering out potential contaminants (Blat) De novo assembly of your reads (Trinity) Reads mapping (Bowtie2) Contigs (i.e transcripts) identification based on: • sequence alignment (Blat, Blast) • protein dominions, profiles (HMMER) • Annotation with GO terms (Blast2GO) Expression quantification: providing TPMs, FPKMs and read counts (RSEM, eXpress). Discover TRUFA with tutorial videos HERE For more information, check out the manual. The service offers an integrated system to perform RNA-seq analysis exploiting computing and storage resources. It is in production and accessible from: https://trufa.ifca.es/web TRUFA has been published in the following paper: Kornobis, Cabellos, Aguilar, Frias-Lopez, Rozas, Marco & Zardoya (2015). TRUFA: A user-friendly web server for de novo RNA-seq analysis using cluster computing. Evolutionary Bioinformatics. 11:97-104. Link to the article
Service provider	LIFEWATCH ERIC SERVICE PROVIDED BY IFCA + MNCN (National Natural History Museum)
Service catalogue	LIFEWATCH SERVICES
Value	Set of integrated software and tools to perform RNA-seq analysis in an easy and friendly way. Access to computing and storage resources for non IT experts.
Current TLR level,	TRUFA (in supercomputer integrated mode) is up and running since

acceptance criteria and validation/verification results	2015 and it is providing service to more than 230 accounts from users around the world, a number that is continuously growing. Correspondingly, more than 2M hours have been freely provided in the last months. The modular architecture separates web service, workflow management, computing and storage, so everything can be updated and switch to other solutions. Currently, a migration to a better scalable computing framework is being developed. **TLR Level: 9** For more information regarding performance and use: Manual: *https://trufa.ifca.es/web/static/trufa_manual.pdf* Kornobis, Cabellos, Aguilar, Frias-Lopez, Rozas, Marco & Zardoya (2015). TRUFA: A user-friendly web server for de novo RNA-seq analysis using cluster computing. Evolutionary Bioinformatics. 11:97-104. Link to the article
Access policy	Free access for academic research under request
Terms of use	Check the manual
User groups and scientific disciplines served	Students, Researchers, Scientists from: Life Sciences, Biodiversity, etc.
Service business model	Free use worldwide.
Service architecture	

Service components

Name of component	Functional description, applicable standards and needed resource capacity (if applicable) e.g. CPU Time, storage capacity etc.	Provider If already appointed
EGI FEDCLOUD	Dynamic deployment of resources. - Long term: Web portal (4 CPUs, 8GB) - Running instances: Up to 256 cores. Up to 32 GB memory.	IFCA/IBERGRID
TRUFA WEB	TRUFA web side. Based on python.	IFCA
INDIGO Orchestrator	Service to dynamically deploy the computing element.	INDIGO
Repositories	Repositories for Docker containers and job script to run the model.	Dockerhub Github

INDIGO IAM	Authentication and Authorization service to access the components.	IAM
Repositories	Repositories for Docker containers and job script to run the model.	Dockerhub Github
OneData	Distributed Storage Service 20 TB Disk	SW: INDIGO Space: IFCA/IBERGRID
FastQC	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/	
Prinseq	http://prinseq.sourceforge.net/	
Cutadapt	https://pypi.python.org/pypi/cutadapt/1.3	
Blat	https://genome.ucsc.edu/FAQ/FAQblat.html	
Trinity	https://github.com/trinityrnaseq/trinityrnaseq/wiki	
Bowtie and Bowtie2	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml	
HMMER	http://hmmer.org/	
BLAST	https://blast.ncbi.nlm.nih.gov/Blast.cgi	

Service integration with generic e-Infrastructures

Integration activity and concerned service components	TRUFA is running on Supercomputer resources with more than 230 users. The following integration activities can contribute to make it more scalable, integrated and dynamic: - Standard use of AAI solutions, working with all the different components. - Use of OneData as distributed storage solution. - Dynamic deployment of running components, using Virtual Machines or Dockers in FedCloud resources. - DevOps oriented deployment of the system.
List of requested service components	EGI FedCloud: Compute, Storage; AARC: AAI. INDIGO: Software (IAM, OneData, Orchestrator).

Infrastructure integration	
Description of infrastructure integration activities	TRUFA will use FedCloud computing resources for running the workflow. A standardized AAI solution will be used (INDIGO IAM) and also a distributed storage system (INDIGO OneData).
Training	
Description of training activities	A set of manuals and Video-tutorials are available at the web site. They will be updated with the new functionalities. Webinars and workshop will be organized as needed.

16- Declic

Service overview	
Thematic Service name	**Declic**
Service description	Declic is a web portal based on Galaxy for metabarcoding and molecular taxonomy (data analysis and machine learning). It provides user friendly access to a set of data analysis and machine learning methods to study the patterns of similarities between molecular based and morphological based diversity in a sample of specimen or individuals, and to establish a dictionary between both. The address for the service is https://galaxy-pgtp.pierroton.inra.fr Documentation is available at https://galaxy-pgtp.pierroton.inra.fr/static/MIAB_doc.pdf
Service provider	PGTB (Genome Transcriptome Facility of Bordeaux - INRA) http://pgtb.cgfb.u-bordeaux.fr/
Service catalogue	This service is part of the Galaxy based catalogue of PGTB https://galaxy-pgtp.pierroton.inra.fr. It is planned to add the service to : • The EGI confluence catalogue (https://confluence.egi.eu/display/EGI/Declic) • EGI CC LW catalogue
Value	The expected benefit is to have a user-friendly access to a series of statistical tools without the necessity to call them through a programming language. The men-machine interface is provided by web-based platform, and filling boxes in forms. This eliminates the entire burden to learn a computing language for biologists who can concentrate on the best use of the tools. This service has been developed to fill the current gap between statistical ecology and biodiversity studies, and connect both research fronts in each of this domain: evolution from natural history to molecular based taxonomy (barcoding and metabarcoding), and evolution from data analysis to machine learning. What has been done for taxonomy can be further extended to other domains in biological sciences.

Current TLR level, acceptance criteria and validation/verification results	The Declic service has been running since 2014. . It is used by several academic partners (five different groups who use it regularly, and twice as many who come and leave from time to time). This service is part of the LifeWatch Competence centre of EGI, for integration as a thematic service in LW Virtual Research Environment. There is no monitoring tool for availability status. Galaxy provides statistics on use, per user. The following published papers have used this service: Rimet F., Trobajo R., Mann D.G., Kermarrec L., **Franc A.,** Domaizon I., Bouchez A., 2014. When is sampling complete? The effects of geographical range and marker choice on perceived diversity in Nitzschia palea (Bacillariophyta).Protist **(165):**245-269. Rimet F., Chaumeil P., Keck F., Kermarrec L., Vasselon V., Kahlert M., Franc A & Bouchez A., 2016. R-Syst::diatom: An open-access and curated barcode database for diatoms and freshwater monitoring. Database: The Journal of Biological Databases and Curation. 2016, 1-21.
Access policy	Policy based service: The service is free and accessible for free by any scientist in an academic research laboratory. Access is granted provided the user provides an institution e-mail address, which is used as a login, and a password is generated for privacy policy. Access can be granted for free o to private research companies upon request (which has been granted once in 2016).
Terms of use	The terms of use as well as technical constraints (like file size) are publicly accessible at https://galaxy-pgtp.pierroton.inra.fr/root Access policy is given here above.
User groups and scientific disciplines served	This service is part of the data management and computing infrastructure of R-Syst network. It is connected to a database. The discipline served is molecular based taxonomy in a precise sense (sensu stricto), and study of biodiversity patterns in a larger sense (sensu lato).
Service business model	The service is part of the activity of the INRA network R-Syst, financially supported by Divisions Ecology of Forests, Grassland and Fresh water, and Plant health and environment of INRA, each year, on basis of a yearly activity report and submission of further plans. The people involved in maintaining the infrastructure, the Galaxy server and the implemented tools have a permanent position at INRA, Bordeaux. Hence the service provided is perennial. Some

<table>
<tr><td></td><td>developments are necessary for connecting the service to EGI cloud and AII. Hiring an engineer is therefore suitable. The service is currently designed on an academic basis. Some connections are curently under constructions with (I) monitoring water quality (European Framework Directive on Water) through molecular based diatom inventories (standard index) and (ii) monitoring of atmospheric allergenic pollens with metabarcoding. Those both connections are under development at National level, and will be fostered at European level with the service provided at European scale.</td></tr>
<tr><td colspan="2">Service architecture</td></tr>
<tr><td>Service components</td><td>
<table>
<tr><th>Name of component</th><th>Functional description, applicable standards and needed resource capacity (if applicable)
e.g. CPU Time, storage capacity etc.</th><th>Provider
If already appointed</th></tr>
<tr><td>Web frontend</td><td>Scientific computation in machine learning (python and standard libraries, numpy, matplotlib, scikit-learn, igraph);
Scientific Linux server; 32 cores, 32 Go RAM</td><td>PGTB</td></tr>
<tr><td>Compute infrastructure</td><td>Virtual machines in the cloud – tested successfully in collaboration with UPV (EGI CC LifeWatch)</td><td>EGI</td></tr>
<tr><td>Data storage and management</td><td>Via a connection with an area in iRODS system</td><td>MCIA, currently, France-Grille next.</td></tr>
</table>
</td></tr>
<tr><td colspan="2">Service integration with generic e-Infrastructures</td></tr>
<tr><td>Integration activity and concerned service components</td><td>Elastic Deployment of virtual machines on EGI cloud, using first prevalidated recipes and ansible roles provided by EC3 middleware, and subsequently designing them.</td></tr>
<tr><td>List of requested service components</td><td>Compute resources.</td></tr>
</table>

Infrastructure integration	
Description of infrastructure integration activities	Some development is needed for full scale deployment (scalability), by dynamic instantiations of virtual machines upon demand. Therefore, our service has been tested within VO EGI FedCloud with EC3 middleware developed by UPV, Valencia. The EC3 service has a development plan to offer pre-validated recipes and Ansible roles in a trusted repository to guarantee that users who use that repository do not incur in any security breach and that the VMs generated are safe according to the EGI standards. The UPV confirms its willingness to support the group from INRA in the development of new configurations and the support to users.
Training	
Description of training activities	A formation on the Galaxy based service has been organized within R-Syst network (the scientific community for molecular based taxonomy). We plan to organize once a year such a formation for a wider audience for users. We plan to organize such a training session on European scale.