

Lifewatch computational services

- ✚ Manage data lifecycle: storage, processing, analysis, preservation...
- ✚ One of the goals of the Lifewatch ICT Core architecture is linking different resources from third party contributors (databases, sensors, software, computing power).
- ✚ Use Case using european infrastructures: EGI, EUDAT



EGI - European Grid Initiative

- ❖ European Grid Initiative - Distributed computational resources.
- ❖ Users must belong to Virtual Organizations (VO).
- ❖ vo.lifewatch.eu available:
 - ❖ <https://ibergrid-voms.ifca.es:8443/voms/vo.lifewatch.eu/register/start.action>
- ❖ Users of a VO needs similar resources.
- ❖ Example: Ugr. Analysis of NDVI in Spain using NASAs Terra Satellite and MODIS sensor.

- ❁ European data: Collaborative Data Infrastructure that allows users to share data within and between communities.
- ❁ Different services:
 - ❁ B2SHARE: Upload and store data. Usually small data sets (citizens or not official data).
 - ❁ B2FIND: Metadata for data stored in EUDAT data centers to make data easy to find.
 - ❁ B2SAFE: Replicate data to selected data centers for storage and do this in a robust, reliable and highly available manner.
 - ❁ B2STAGE: Data to HPC facilities.
- ❁ Other services: Authentication, PIDs, Training...

B2FIND

Aggregated EUDAT metadata domain.
Data inventory



B2SAFE

Data curation and
access optimization



B2STAGE

Dynamic replication
to HPC workspace
for processing



B2SHARE

Researcher data
store (simple
upload, share and
access)



AAI

Network of trust
among
authentication
and authorization
actors

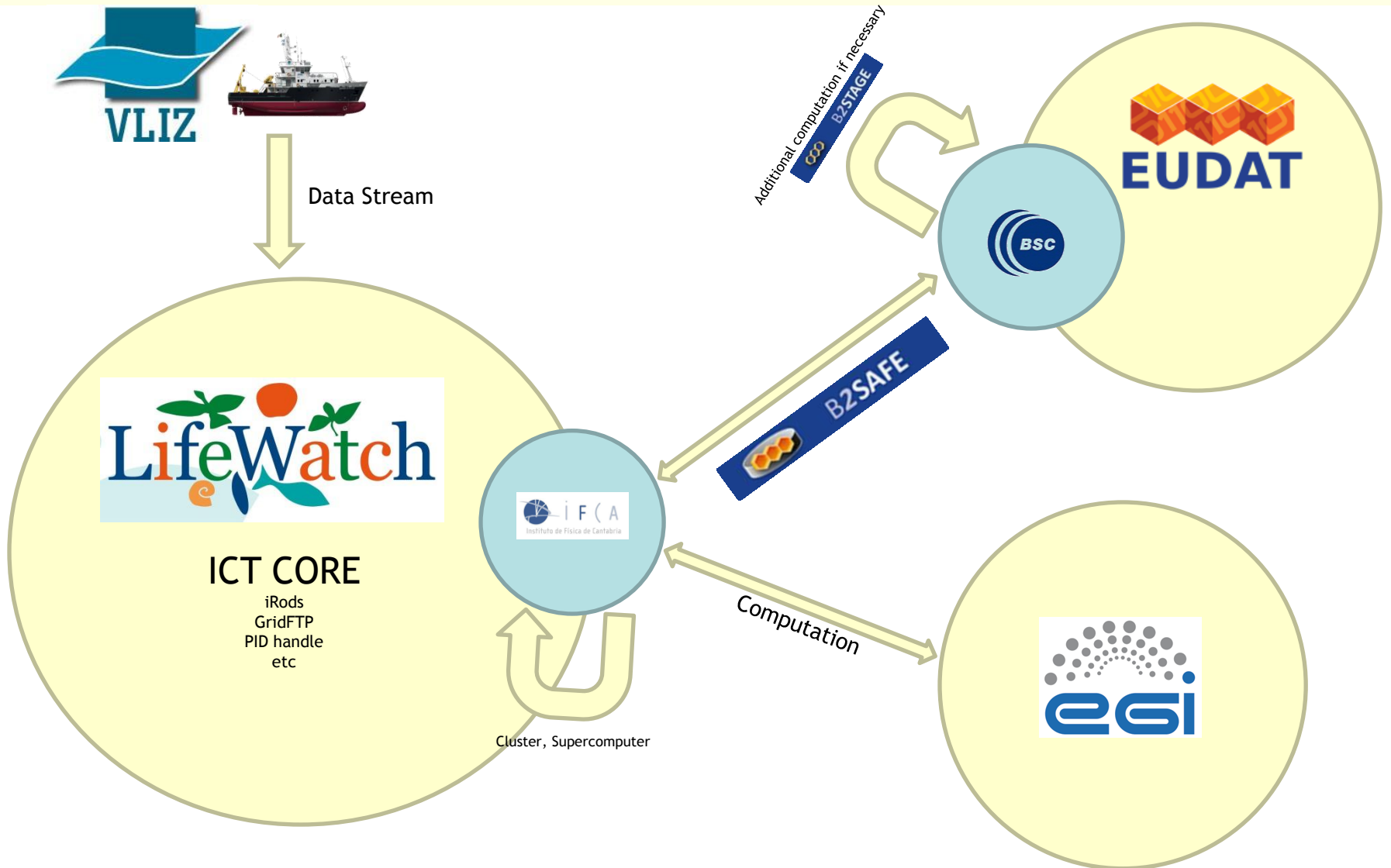


PID

Identity
Integrity
Authenticity
Locations



Use case



Web Interface Pipeline: TRUFA

- ⊕ One approach can be a web interface to guide users in data management and processing.
- ⊕ TRUFA is a pipeline that allows users to process RNA sequences.
- ⊕ Users upload the files, select different options and run the job(s).
- ⊕ File manager to handle input and outputs.
- ⊕ Perfect for non-experts users in informatics

Web Interface Pipeline: TRUFA

The screenshot shows the TRUFA 0.7.0 home page. The navigation bar includes links for Home, Start a Job, File Manager, How to, FAQ, About, Bugs, and a user profile for 'agularf' with a Logout button. The main content area features a 'Welcome to TRUFA' section with a brief description and a 'Help' link. To the right is a 'File Upload' section with a file input field, a 'Browse' button, an 'Upload' button, and a dropdown menu for 'Choose an input format:'. Below these are two columns: 'Files' listing news.txt, reads_left_dup.log, Trinity.fasta, reads_left.fq.tar.gz, reads_left_renamed.fq, and reads_righttt.fq; and 'Jobs' listing job 1 and job 2.

The screenshot displays the 'RNA-seq steps:' configuration page. It includes a note: 'You can perform RNA-seq steps independently or sequentially depending on the boxes you check in each step tabs:'. The steps are: 1. Cleaning step: Pre-cleaning quality control (FastQC), Removing adapters (Cutadapt), Prinseq (Duplicated reads, Quality Trimming), and BLAT against potential contaminants (Univec hits, E. coli hits, S. cerevisiae hits). A 'Nucleotide db' link is also present.

The screenshot shows the 'FAQ:' section. It addresses 'What is TRUFA's purpose?', 'What type of input data can I use?' (listing Illumina reads and assembled transcripts), 'How much can last a submitted job to finish?' (dependent on input size and server load), and 'How do I get the results of an analysis?' (via the File Manager). It also lists the output folders: CLEANING, ASSEMBLY_MAPPING, IDENTIFICATION, EXPRESSION, and STAT.

The screenshot shows the 'File Manager' interface. The current folder is '/agularf/data/'. The left sidebar shows a tree view with folders like 'admin', 'agularf', 'data', 'jobs', 'kor-nobis', 'rotas', 'zardoya', and 'zardoya2'. The main area displays three files: reads_left.fq, reads_left_dup.log, and reads_right.fq. A search bar is located at the bottom.