**MAchinE Learning for Scalable meTeoROlogy and climate**


# Web Portal including Software Collaboration Platform

Daniel Thiemert, Jan Mirus, Peter Dueben

www.maelstrom-eurohpc.eu


# D4.4 Web Portal including Software Collaboration Platform


| | |
|---|---|
| **Author(s):** | Daniel Thiemert (ECMWF), Jan Mirus (for 4cast), Peter Dueben (ECMWF) |


| | |
|---|---|
| **Dissemination Level:** | Public |
| **Date:** | 29/07/2021 |
| **Version:** | 1.0 |
| **Contractual Delivery Date:** | 31/07/2021 |
| **Work Package/ Task:** | WP4/ T4.3 |
| **Document Owner:** | ECMWF |
| **Contributors:** | All Partners |
| **Status:** | Final |

# MAELSTROM

# Machine Learning for Scalable Meteorology and Climate

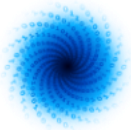
**Research and Innovation Action (RIA)**
**H2020-JTI-EuroHPC-2019-1: Towards Extreme Scale Technologies and Applications**

**Project Coordinator:** Dr Peter Dueben (ECMWF)
**Project Start Date:** 01/04/2021
**Project Duration:** 36 months

**Published by the MAELSTROM Consortium**

**Contact:**
ECMWF, Shinfield Park, Reading, RG2 9AX, United Kingdom
Peter.Dueben@ecmwf.int

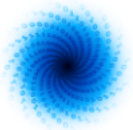
The MAELSTROM project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 955513. The JU receives support from the European Union's Horizon 2020 research and innovation programme and United Kingdom, Germany, Italy, Luxembourg, Switzerland, Norway

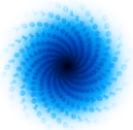


# Contents



# Figures





# 1 Executive Summary

The MAELSTROM web portal is the main dissemination and interaction tool of the MAELSTROM project, both in relation to external parties and project partners. The domain website (available via www.maelstrom-eurohpc.eu), provides detailed information on the project (including news and events, objectives, impact, etc.). The web portal will be regularly maintained, with new content being added as the project progresses.

With development at the core of the MAELSTROM project, a collaborative software development environment is essential for its success. Such environments require tools for bug tracking, software repositories, and wikis.

This deliverable therefore also presents the software collaboration platform for the MAELSTROM project, consisting of JIRA (Bug Tracking), Confluence (Collaborative Wiki), and Bitbucket (Software Repository).

To access the platform, the following details are required:

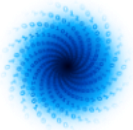
**Confluence:** https://confluence.ecmwf.int/display/MLFET/

**JIRA:** https://jira.ecmwf.int/projects/MAEL/summary

**Bitbucket**: https://git.ecmwf.int/projects/MLFET

(Users are required to have a login for www.ecmwf.int to be able to access the data. This can be requested via the coordinator.)

Thus, this document provides an introduction to the web portal as an accompanying document to the actual online portals, and, presents a brief introduction to each of the software collaboration tools, and provides links to extensive online user manuals.



# 2 Introduction

## 2.1 About MAELSTROM

To develop Europe's computer architecture of the future, MAELSTROM will co-design bespoke compute system designs for optimal application performance and energy efficiency, a software framework to optimise usability and training efficiency for machine learning at scale, and large-scale machine learning applications for the domain of weather and climate science.

The MAELSTROM compute system designs will benchmark the applications across a range of computing systems regarding energy consumption, time-to-solution, numerical precision and solution accuracy. Customised compute systems will be designed that are optimised for application needs to strengthen Europe's high-performance computing portfolio and to pull recent hardware developments, driven by general machine learning applications, toward needs of weather and climate applications.

The MAELSTROM software framework will enable scientists to apply and compare machine learning tools and libraries efficiently across a wide range of computer systems. A user interface will link application developers with compute system designers, and automated benchmarking and error detection of machine learning solutions will be performed during the development phase. Tools will be published as open source.

The MAELSTROM machine learning applications will cover all important components of the workflow of weather and climate predictions including the processing of observations, the assimilation of observations to generate initial and reference conditions, model simulations, as well as post-processing of model data and the development of forecast products. For each application, benchmark datasets with up to 10 terabytes of data will be published online for training and machine learning tool-developments at the scale of the fastest supercomputers in the world. MAELSTROM machine learning solutions will serve as blueprint for a wide range of machine learning applications on supercomputers in the future.

## 2.2 Scope of this deliverable

### 2.2.1 Objectives of this deliverable

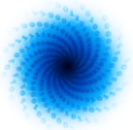
D4.4 describes the MAELSTROM Project website as well as the project-internal software collaboration platform.

### 2.2.2 Work performed in this deliverable

The website www.maelstrom-eurohpc.eu was implemented, and the internal software collaboration platform was setup using the Atlassian tools Jira, Confluence and Bitbucket.

### 2.2.3 Deviations and counter measures

No deviations have been encountered.



# 3 The MAELSTROM Website

The domain website is reachable via www.maelstrom-eurohpc.eu. It is the main dissemination tool for the project and provides both external as well as project-internal sections. The website is targeting a range of different audiences and aims to be appealing for the general public (to get a general idea what MAELSTROM is about and to realise that MAELSTROM is a vivid and active project), interested machine learning scientists and engineers (who want to know whether MAELSTROM tools could help them for their day-to-day work, or whether they can find a solution that MAELSTROM may provide for their machine learning or high-performance computing problem), domain scientists working in weather and climate prediction (who want to learn more about the capabilities of machine learning, and industry and in particular hardware vendors (who want to learn more about the use of MAELSTROM machine learning problems for high-performance computing benchmarks). The site navigation is given below:

- MAELSTROM Home
  - Our Mission
    - ML apps & data
    - ML dev tools
    - HPC systems
  - We are MAELSTROM
    - Consortium Partners
    - MAELSTROM Team
    - Project Funders
  - News
  - Follow our progress
    - Timeline
    - Events
  - Products
    - ML apps & data
    - ML dev tools
    - HPC systems
  - Documentation
  - Contact Us

Site navigation is enabled through the top navigation bar. The site navigation aims to provide an overview page with more generic information and deeper structure to provide additional information that can be more technical.

## 3.1 Website sections

### 3.1.1 Landing Page

The landing page contains high-level information with links to further details, as well as latest news items and consortium partners.


*Figure 1: MAELSTROM Landing Page*

### 3.1.2 Our mission

This section introduces the mission of the project, with respect to weather and climate forecasting, machine learning, and high performance computing.


*Figure 2: Our mission*

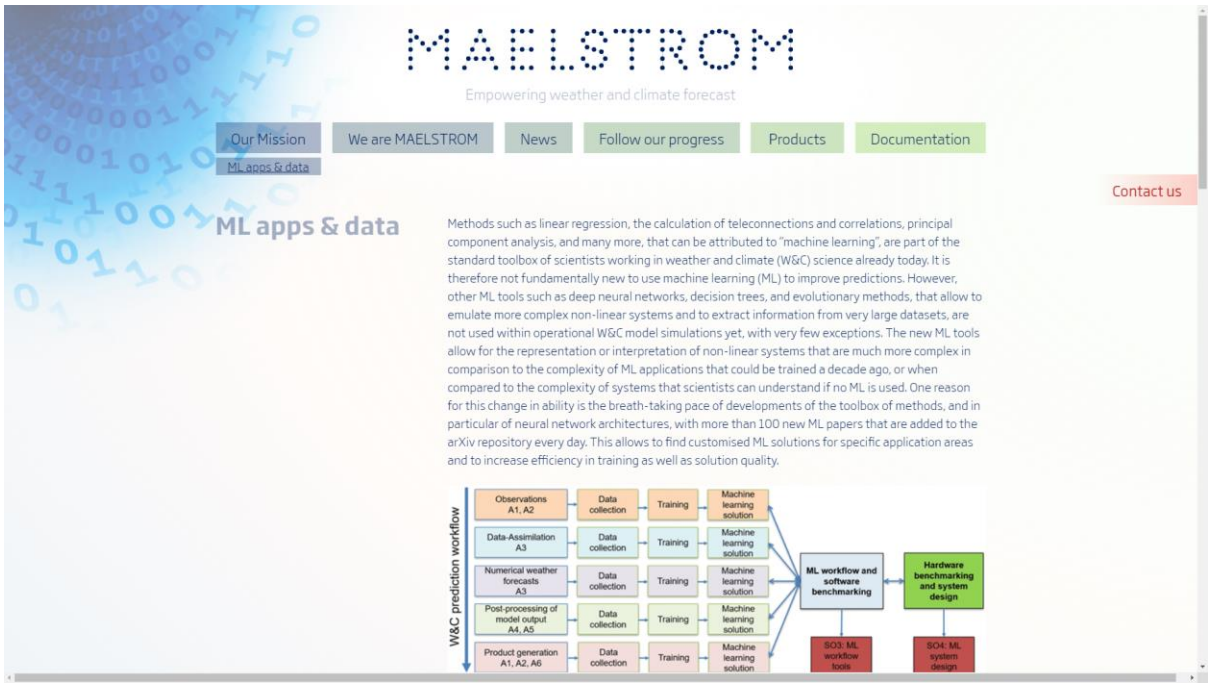



*Figure 3: ML apps & data*


*Figure 4: ML dev tools*


*Figure 5: HPC systems*

### 3.1.3 We are MAELSTROM

This section introduces the consortium and team behind the project.


*Figure 6: Consortium Partners*

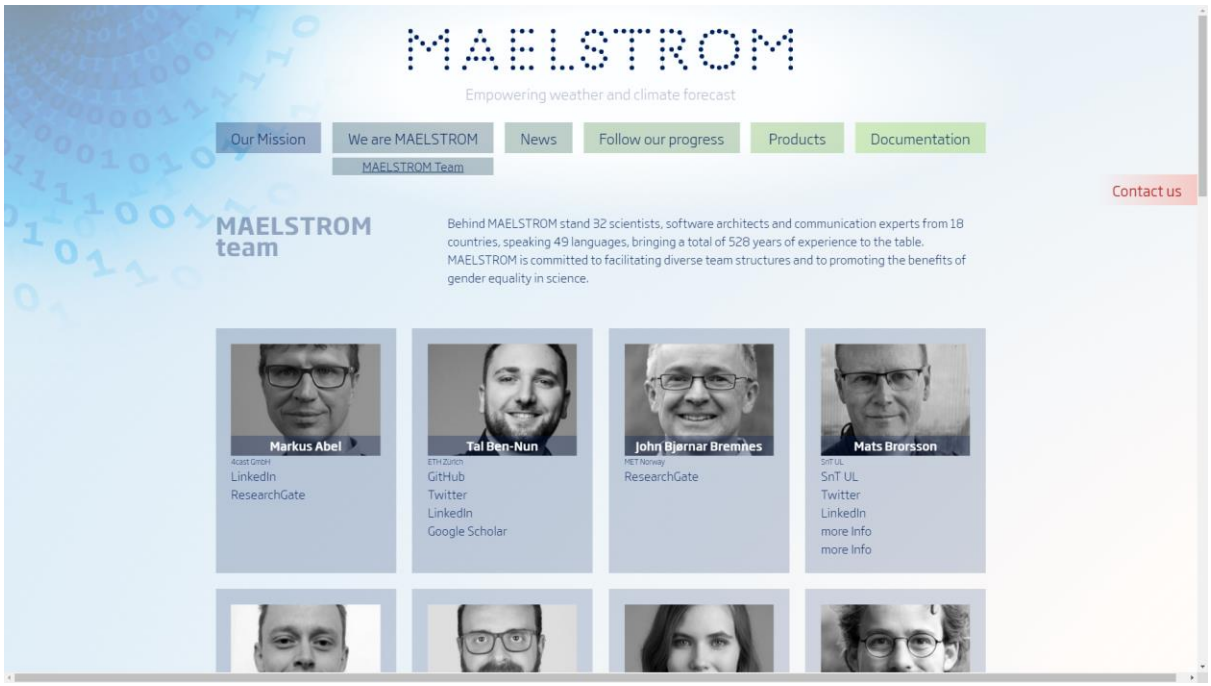



*Figure 7: MAELSTROM Team*


*Figure 8: Project Funders*

### 3.1.4 News

This section provides the news items, the latest of which are also shown on the landing page.

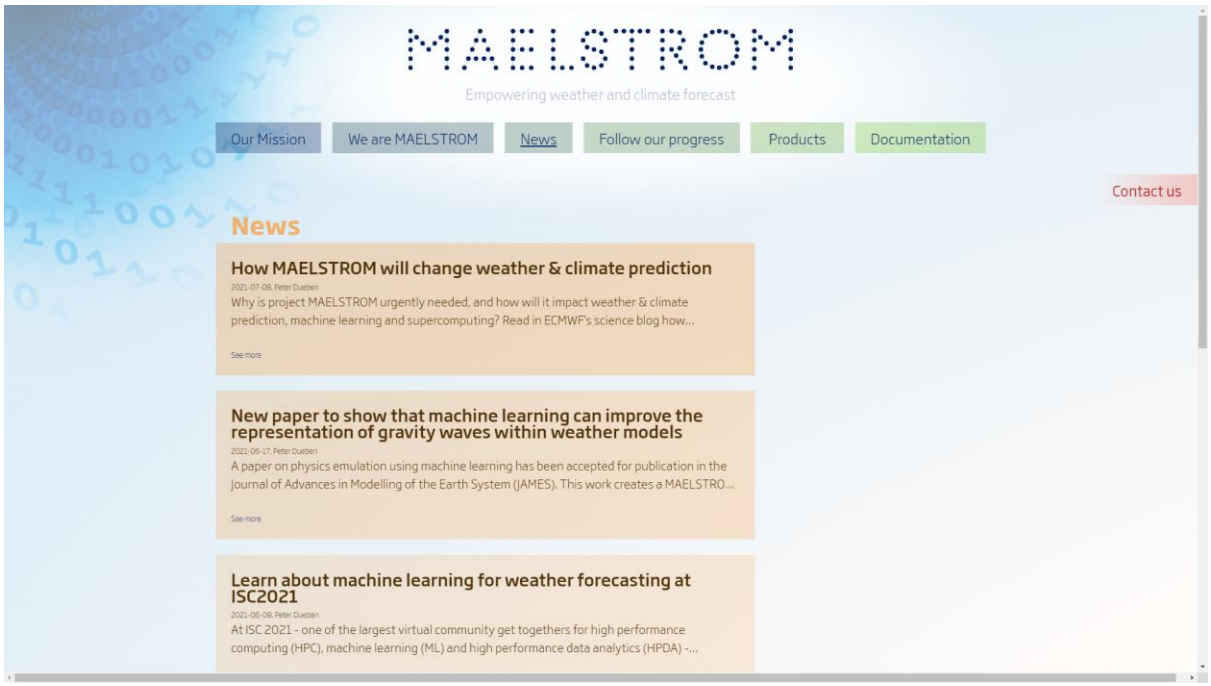



*Figure 9: News*

### 3.1.5 Follow our progress

This section provides both an overview of the progress with a timeline, as well as upcoming events.

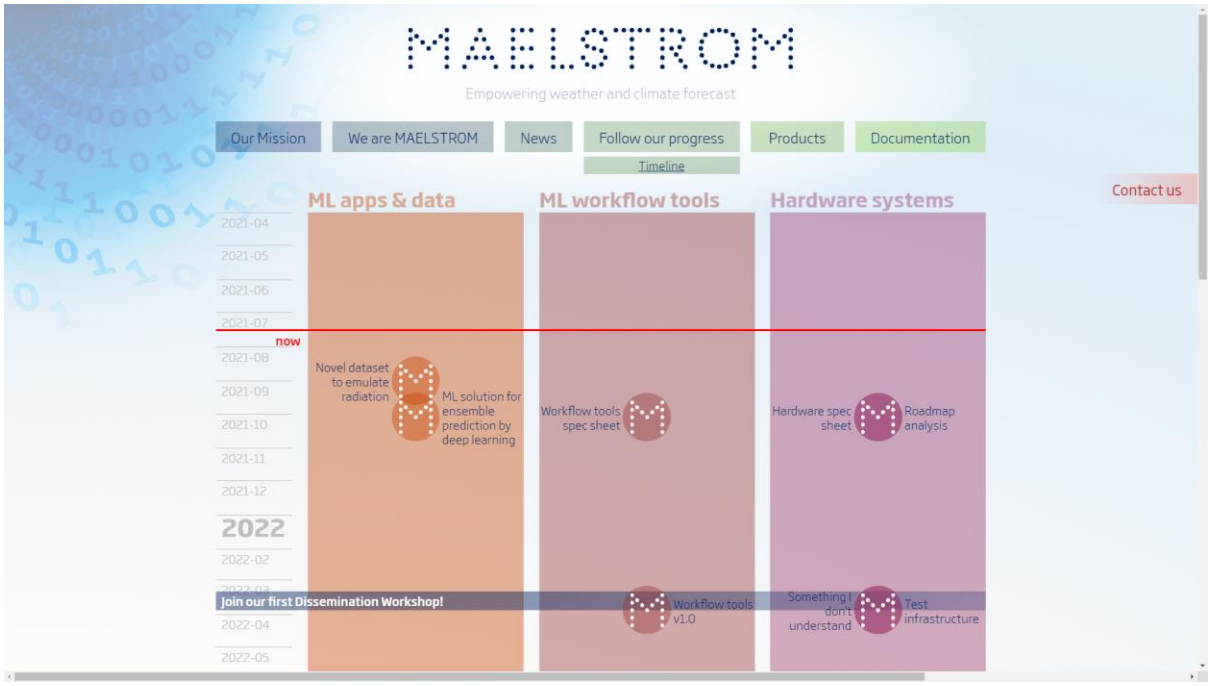



*Figure 10: Timeline*


*Figure 11: Events*

### 3.1.6 Products

This section of the website introduces the products (i.e. tangible outcomes) of the project. Links to downloads will be provided here.


*Figure 12: Products*

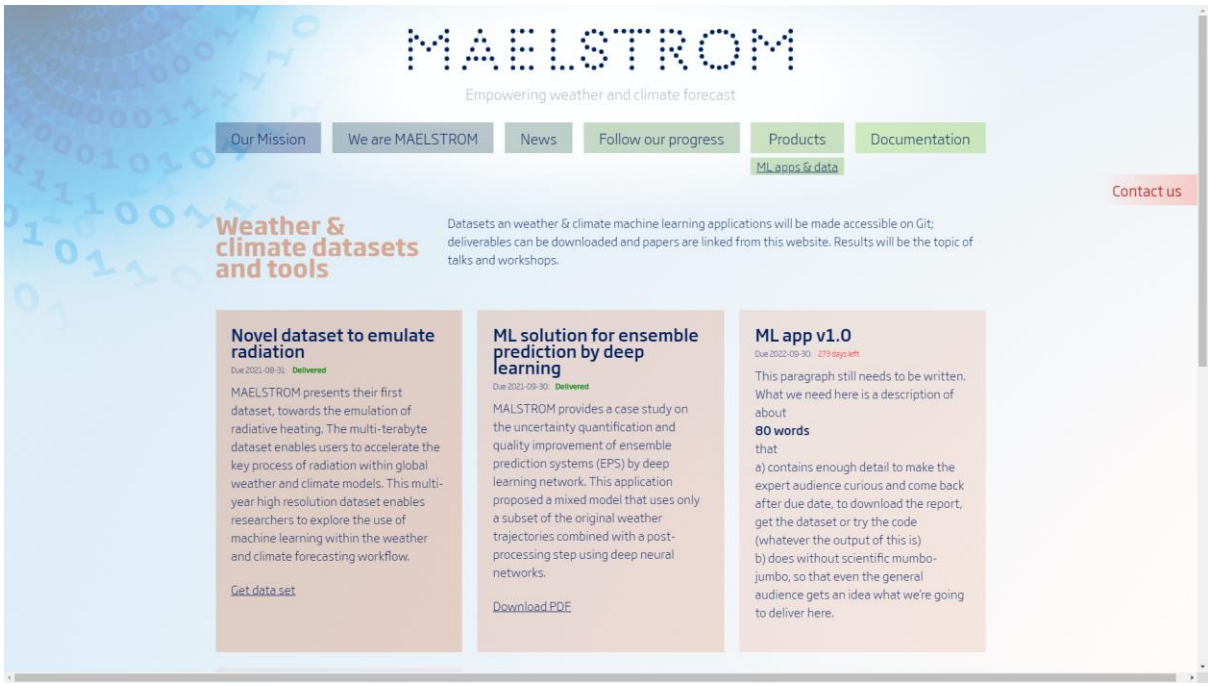



*Figure 13: Products - Weather and climate dataset and tools*

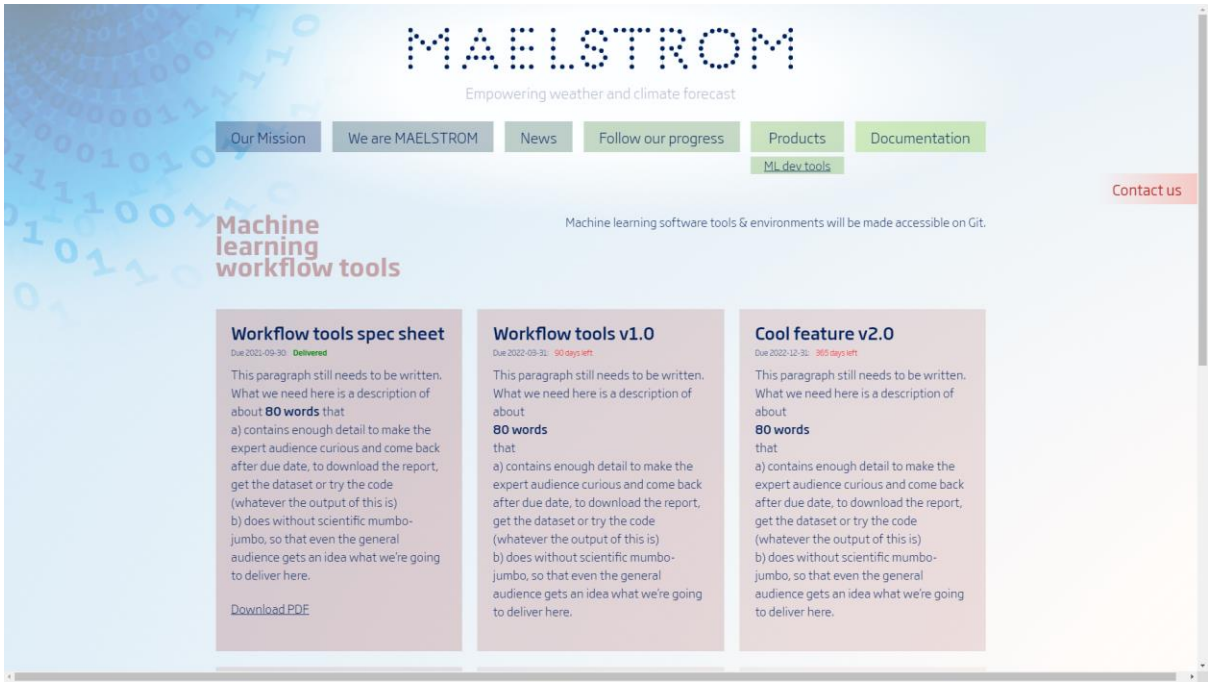



*Figure 14: Products - Machine learning workflow tools*

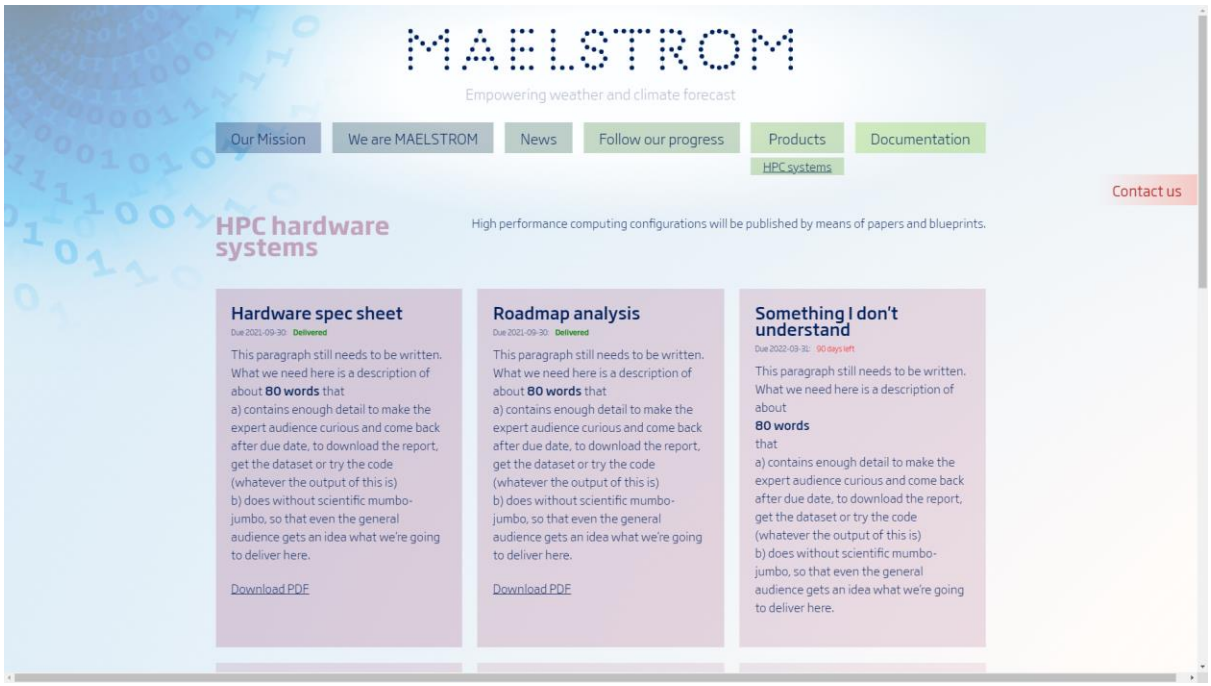
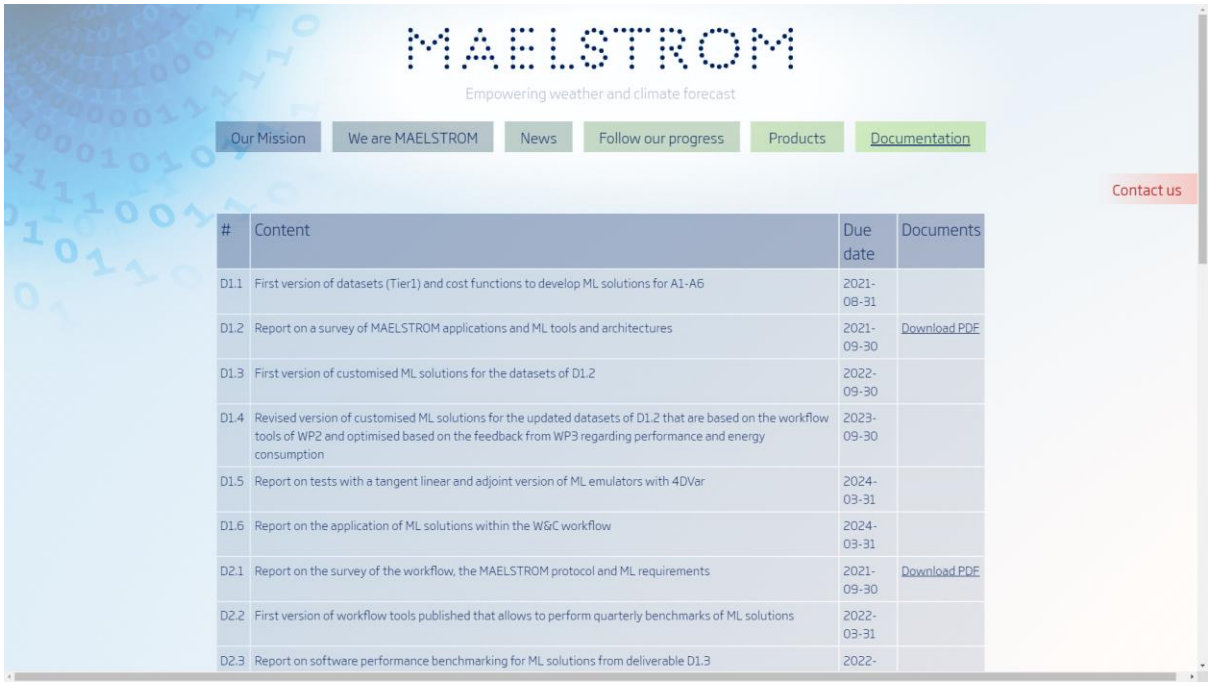



*Figure 15: Products - HPC hardware systems*

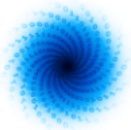
### 3.1.7 Documentation

This section provides and overview and links to downloads of project documentation, including reports, papers, and presentations.




| # | Content | Due date | Documents |
|---|---|---|---|
| D1.1 | First version of datasets (Tier1) and cost functions to develop ML solutions for A1-A6 | 2021-08-31 | |
| D1.2 | Report on a survey of MAELSTROM applications and ML tools and architectures | 2021-09-30 | Download PDF |
| D1.3 | First version of customised ML solutions for the datasets of D1.2 | 2022-09-30 | |
| D1.4 | Revised version of customised ML solutions for the updated datasets of D1.2 that are based on the workflow tools of WP2 and optimised based on the feedback from WP3 regarding performance and energy consumption | 2023-09-30 | |
| D1.5 | Report on tests with a tangent linear and adjoint version of ML emulators with 4DVar | 2024-03-31 | |
| D1.6 | Report on the application of ML solutions within the W&C workflow | 2024-03-31 | |
| D2.1 | Report on the survey of the workflow, the MAELSTROM protocol and ML requirements | 2021-09-30 | Download PDF |
| D2.2 | First version of workflow tools published that allows to perform quarterly benchmarks of ML solutions | 2022-03-31 | |
| D2.3 | Report on software performance benchmarking for ML solutions from deliverable D1.3 | 2022- | |

*Figure 16: Documentation*



# 4 Software Collaboration Platform

The software collaboration platform consists of the Atlassian tools Confluence, Jira and Bitbucket.

To access the platform, the following details are required:

**Confluence:** https://confluence.ecmwf.int/display/MLFET/

**JIRA:** https://jira.ecmwf.int/projects/MAEL/summary

**Bitbucket**: https://git.ecmwf.int/projects/MLFET

## 4.1 Confluence

### 4.1.1 Brief Introduction

Confluence is the Atlassian wiki-style collaboration tool that allows easy creation of pages, discussions, as well as integration with JIRA (i.e. linking to JIRA issues). Confluence will mainly be used for internal communication within the project but across all partners.

#### *4.1.1.1 Creating a new page*

Creating a new (sub-) page is done via the “Create” button, as highlighted in Figure 17.

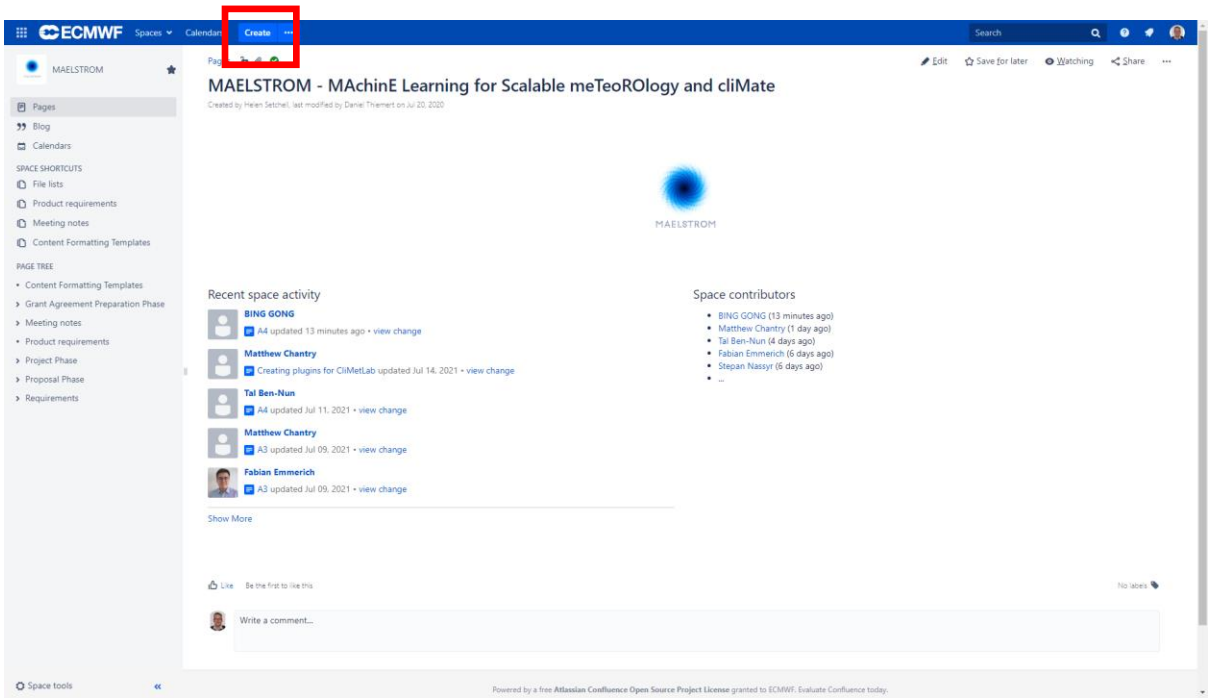



*Figure 17: Creating a new Confluence Page*

This opens up a new blank page.

#### *4.1.1.2 Editing existing pages*

Existing pages can be edited by all project members, through the “Edit” button as highlighted in Figure 18.

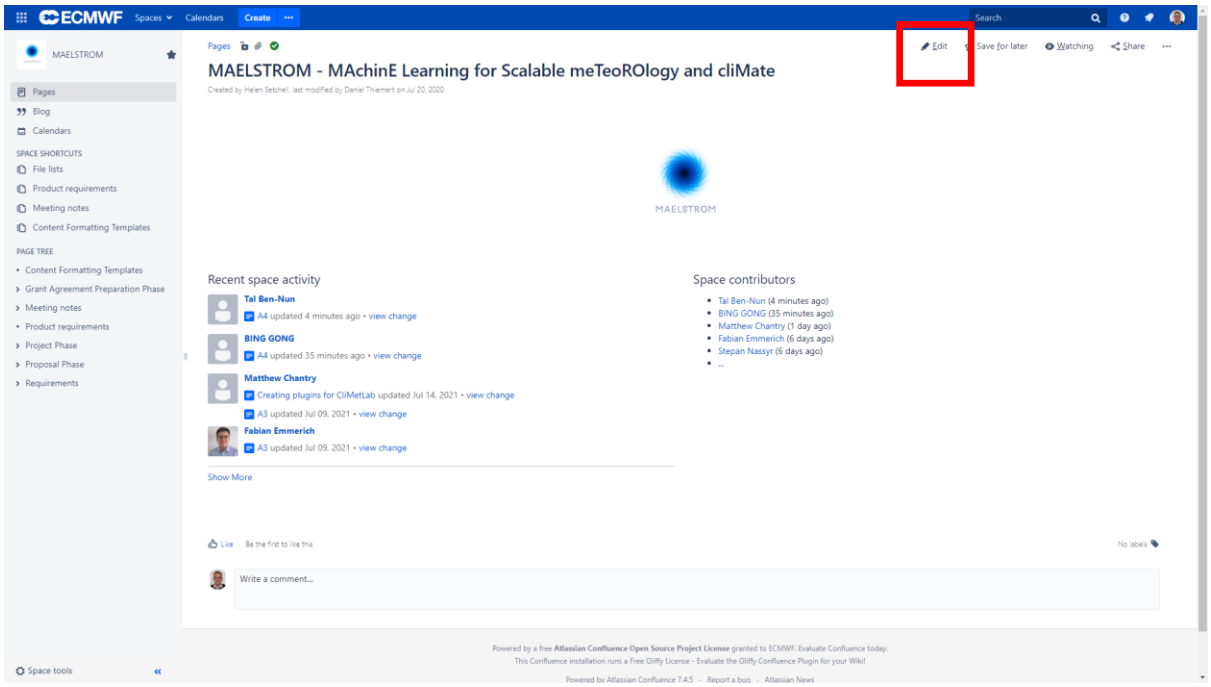



*Figure 18: Editing a new Confluence Page*

Clicking on the "Edit" button will open the edit view of the Confluence page. This allows a "WYSIWYG[1]"-style editing, as shown inFigure 18, with a tool bar for common formatting needs. After editing, the page can be saved by clicking the "Update" button.

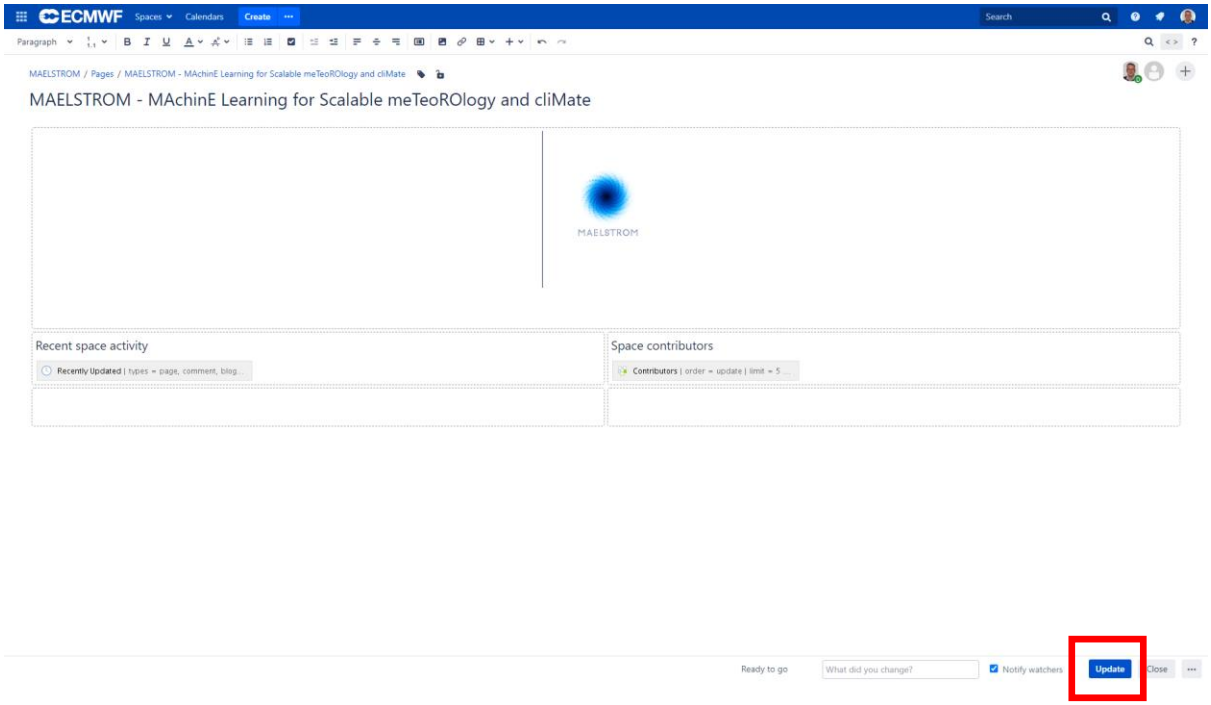
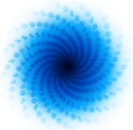



*Figure 19: Edit view*

[1] WYSIWYG - What You See Is What You Get



#### *4.1.1.3 Further reading*

The complete user manual for Confluence is provided online by Atlassian and can be found at:

https://confluence.atlassian.com/doc/confluence-server-documentation-135922.html

### 4.1.2 Access Details

The MAELSTROM confluence space is available at:

https://confluence.ecmwf.int/display/MLFET/.

To gain access to the confluence, users will need to have registered first at

https://apps.ecmwf.int/registration/

and have sent their username to daniel.thiemert@ecmwf.int. Users will then be given the rights to access the space. Once this is done, users should log in at www.ecmwf.int with their registration details and can then navigate to the above-mentioned confluence space.

## 4.2 JIRA

## 4.3 Brief Introduction

Atlassian's JIRA is an Issue Tracking Platform that integrates closely with its other tools Confluence and Bitbucket, and thus forms a powerful collaboration platform that is being deployed in MAELSTROM. JIRA allows for the creation and tracking of issues (e.g. bugs identified in software, new feature requests, etc.), which can be closely linked to existing source code. JIRA will mainly be used as internal tool to concert the efforts of different partners across the MAELSTROM project.

#### *4.3.1.1 Creation of a JIRA Issue*

To create a new Jira Issue, click on the "Create" button in the MAELSTROM Jira project, as highlighted inFigure 20. This will open a pop-up window where details such as issue type (new feature, task, improvement, epic, story), priority, as well as assignee can be provided.

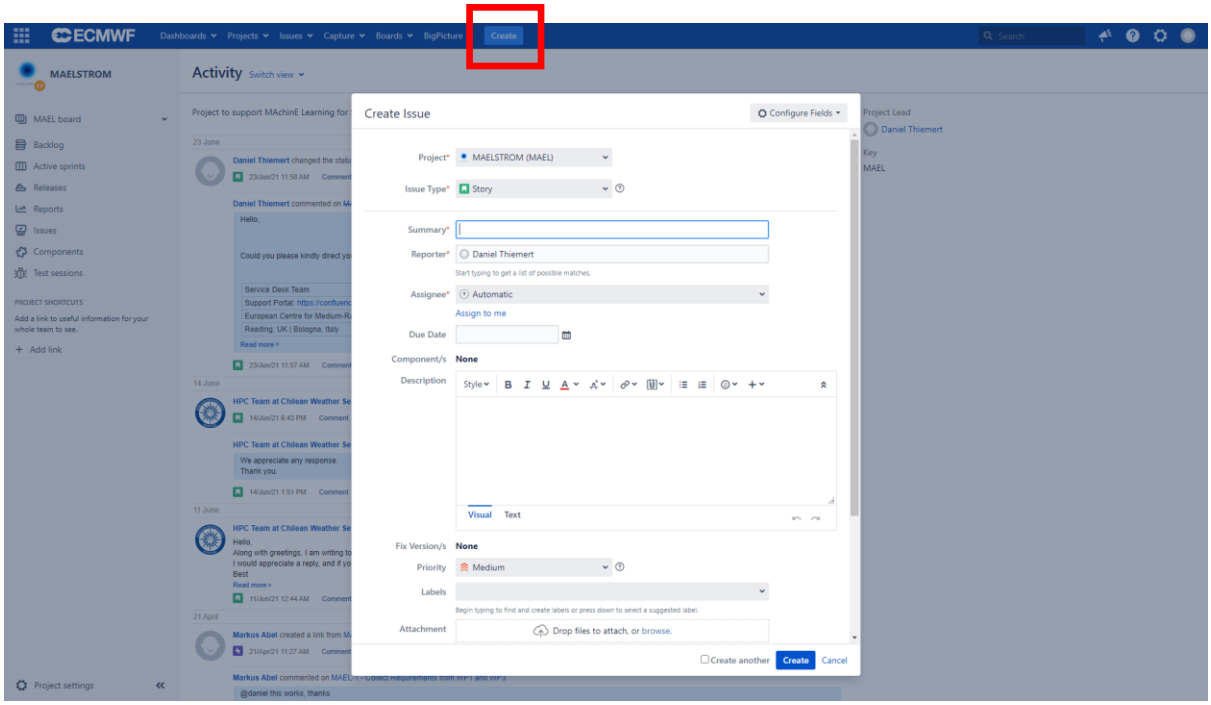



*Figure 20: Creation of a JIRA Issue*

### 4.3.1.2 Creating an Issue-specific branch

Due to the tight integration with Bitbucket, it is possible to create a code branch related to each issue which can, after the issue-specific developments, be merged back into the main code.

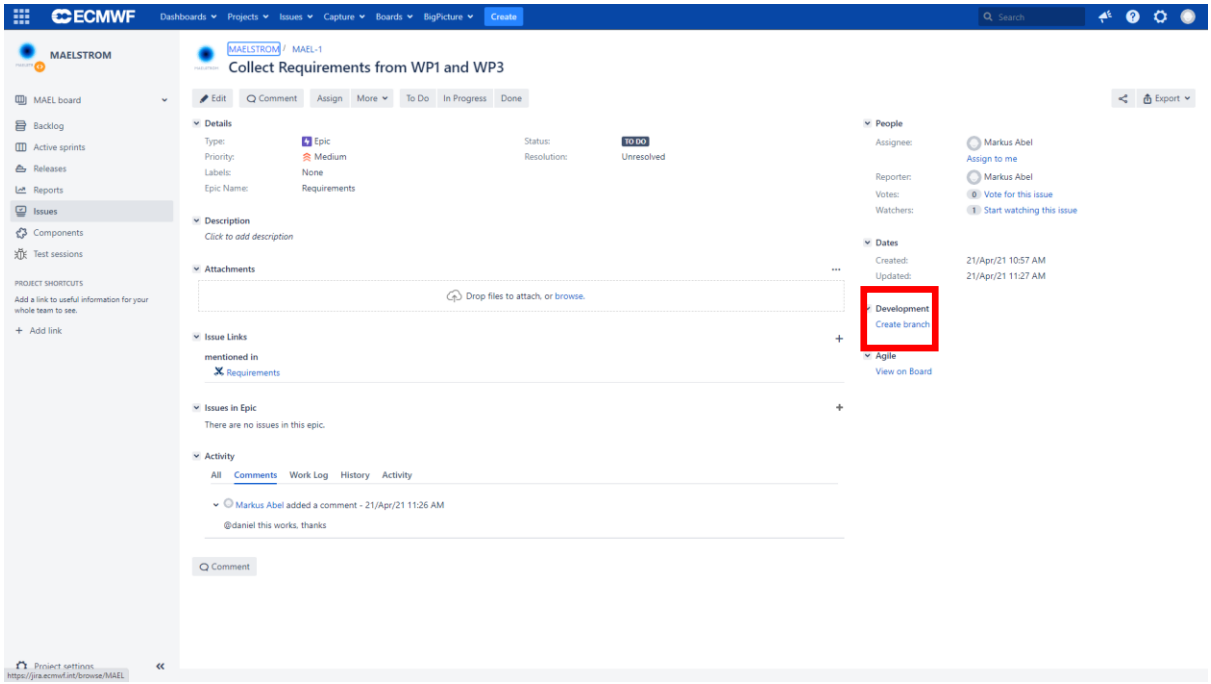
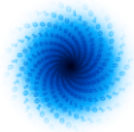



*Figure 21: Creation of an Issue-specific code branch*



#### 4.3.1.3 Modifying Issues

In the Jira interface, issues can be modified, commented on, and the status of the issue can be changed. The interface is shown in Figure 22.

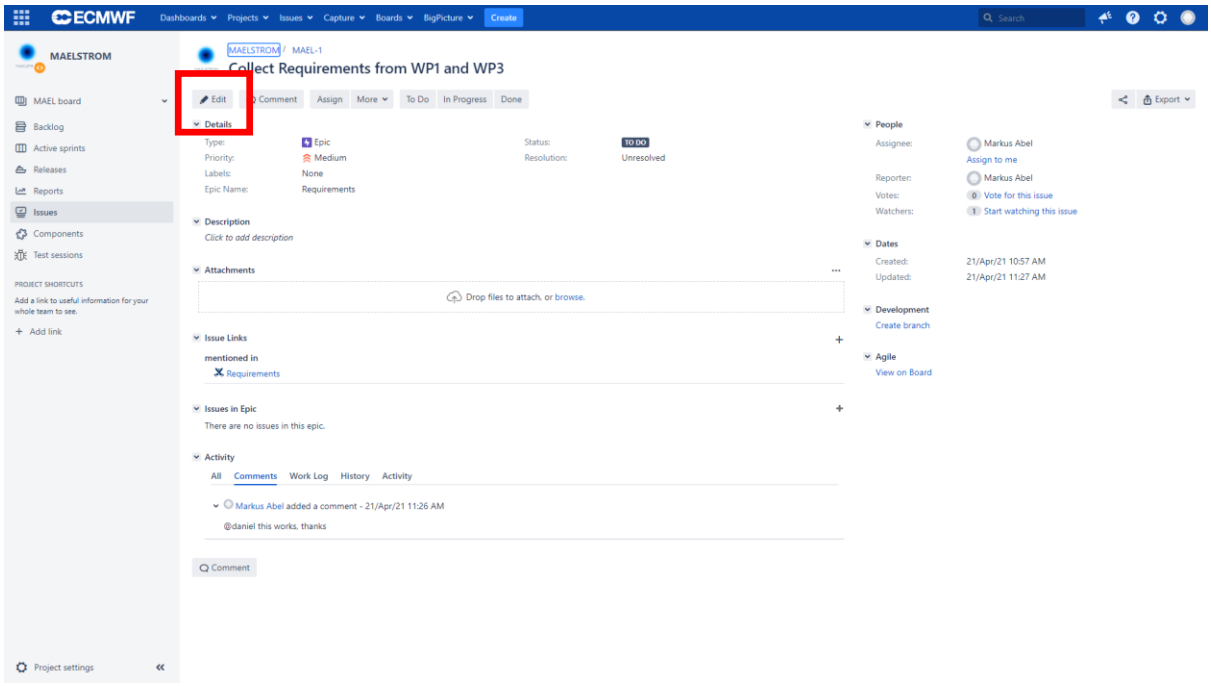



*Figure 22: Modifying issues*

#### 4.3.1.4 Further reading

The complete user manual for JIRA is provided online by Atlassian and can be found at:

https://confluence.atlassian.com/jiracoreserver073/jira-core-server-7-3-documentation-861255603.html.

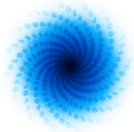
### 4.3.2 Access Details

JIRA is accessible via https://jira.ecmwf.int/projects/MAEL/summary. Users will first have to login via www.ecmwf.int (using the same details as for Confluence - see section 4.1.2) and can then navigate to the JIRA page.

## 4.4 Bitbucket

### 4.4.1 Brief Introduction

Bitbucket is Atlassian's Git hosting service. It offers a web-interface similar to the publicly available Bitbucket (https://bitbucket.org). Through Bitbucket, software repositories can be managed using the distributed version control system Git (https://git-scm.com). Bitbucket will be used for internal code developments within the project. However, Bitbucket will also be used to publish code to the general audience.



#### 4.4.1.1 Cloning a repository

After clicking on a repository, the repository can be cloned by typing in the terminal the command "git clone" followed by the URL given by clicking on Clone/HTTP (see Figure 23).

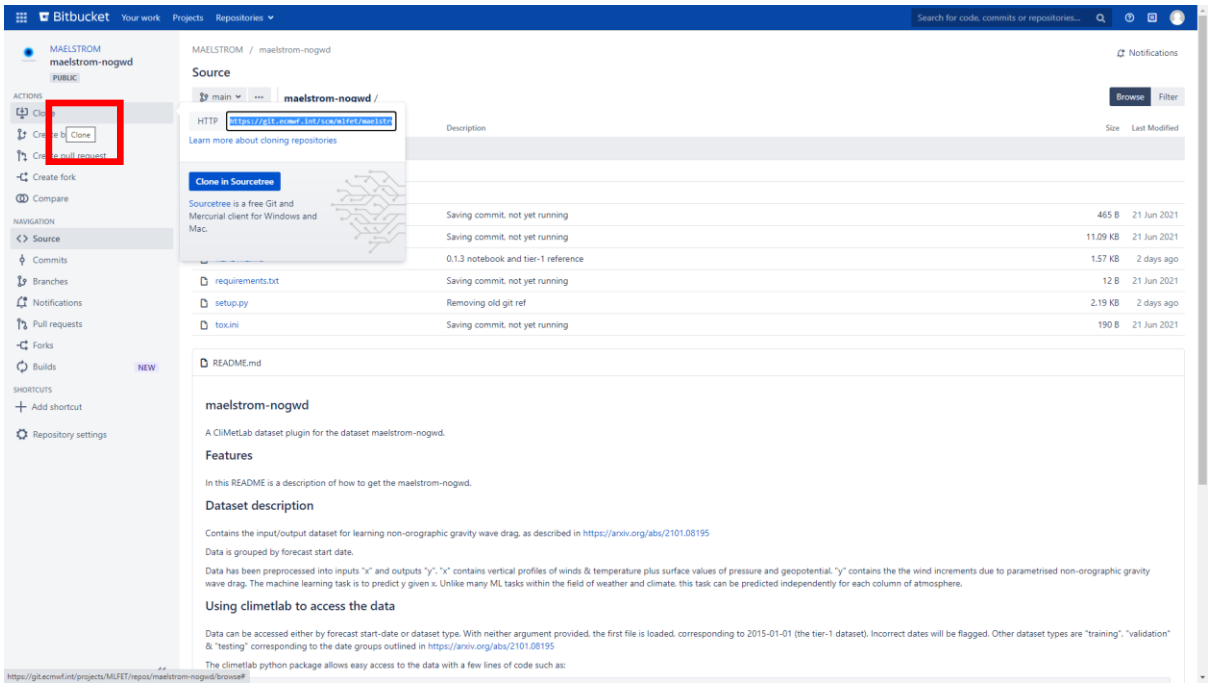



*Figure 23: Cloning Repositories*

#### 4.4.1.2 Creating a new feature branch

To create a new feature branch, browse to the repository and fill in the "Create branch" template according to following figure by clicking on Actions/Create branch (Figure 24).

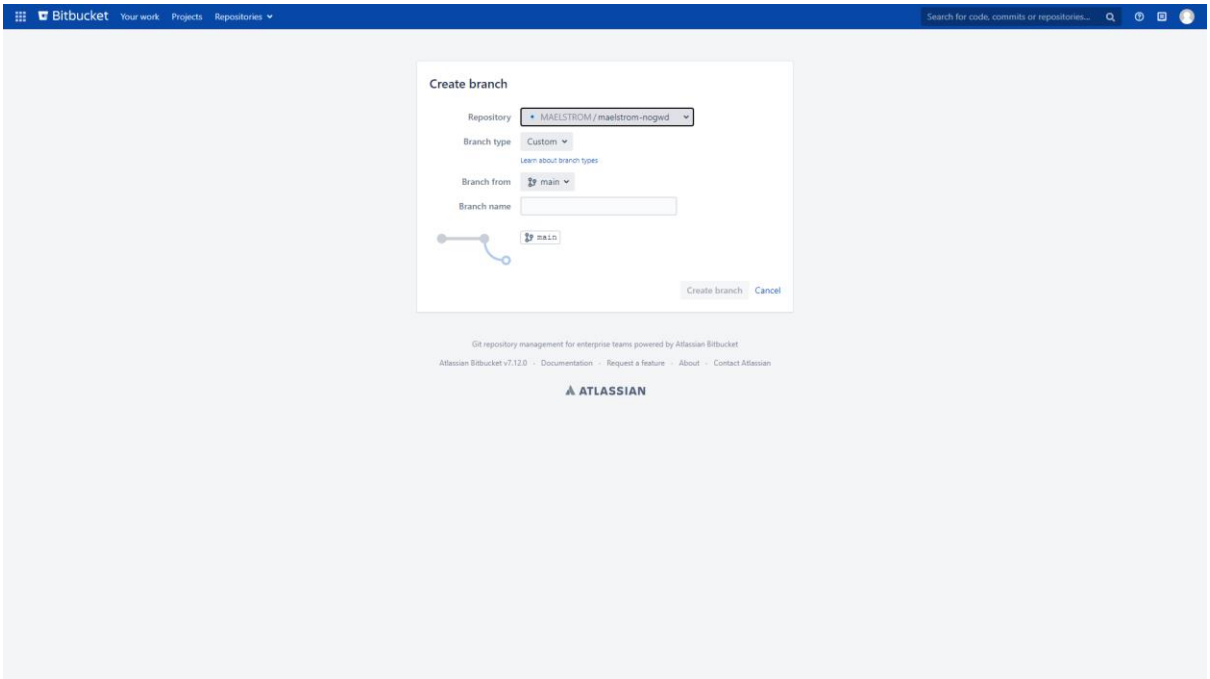



*Figure 24: Creating a new branch*

In the terminal, the new branch can be checked out by typing “git checkout feature/new-improvements”.

#### *4.4.1.3 Further reading*

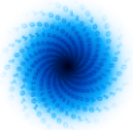
Further reading on using Bitbucket can be found at:

https://confluence.atlassian.com/bitbucketserver0510/bitbucket-server-documentation-951390343.html.

Additionally, links to tutorials on how to use Git are available at this page.

### 4.4.2 Access Details

All repositories required to build the dwarfs provided by ECMWF are available at https://git.ecmwf.int/projects/MLFET. Users will first have to login via www.ecmwf.int (using the same details as for Confluence - see section 4.1.2) and can then navigate to the Bitbucket page.



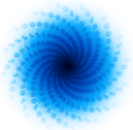
# 5 Conclusion

This document, D4.4, provides a high-level description of the MAELSTROM web portal. It presents details on the structure of the web portal. The sections available in the portal, their content and their use by external and internal users are described. The web portal (accessible via www.maelstrom-eurohpc.eu) is to be updated regularly, both throughout the lifetime of the project and thereafter. It contains information on objectives, news and events, publications (including public deliverables), amongst others.

The document also provides the software collaboration environment for the MAELSTROM project, consisting of three tools:

1. JIRA - a bug tracking tool
2. Bitbucket - a software repository
3. Confluence - a wiki

This deliverable has provided an introduction to each of the tools, including a brief user manual and access details, whilst directing the reader to extensive online manuals. This now provides the consortium with the requisite means to further the implementation of the software for MAELSTROM project.



## Document History

| Version | Author(s) | Date | Changes |
|---|---|---|---|
| **0.1** | Daniel Thiemert (ECMWF), Jan Mirus (for 4cast), Peter Dueben (ECMWF) | 16/07/2021 | Version for Review |
| **1.0** | Daniel Thiemert (ECMWF), Jan Mirus (for 4cast), Peter Dueben (ECMWF) | 25/07/2021 | Revised Version after Review |

## Internal Review History

| Internal Reviewers | Date | Comments |
|---|---|---|
| **Mats Brorsson (Uni Luxembourg)** | 20/07/2019 | Minor changes to motivation and write-up of the deliverable |
| **Greta Denisenko (4cast)** | 22/07/2019 | Minor changes to text and structure |

## Estimated Effort Contribution per Partner

| Partner | Effort |
|---|---|
| **ECMWF** | 0.5 |
| **4cast** | 0.7 |
| | |
| | |
| **Total** | **1.2** |