



developmentSEED

Use Case Descriptor

Global Fish Tracking System



FOR	ESA and Starion
BY	Development Seed

VERSION	1.0
DATE	May 3, 2024



Introduction

This document describes the Global Fish Tracking System (GFTS) use case for DestinE Platform. The subject area of the GFTS use case is in marine biology, the project is performing fish track reconstruction within biologging science and estimation of future fish habitat conditions based on sea temperature projections from the Climate Adaptation Digital Twin.

This document will be continuously updated as we progress in the use case development. Especially the sections around integration with DestinE Platform services and the use of DestinE Digital Twin data will evolve as we advance in implementing the original plan in a collaborative co-creation approach.

Actors

There are three main actors in this use case. The three actors have complementary capabilities that will ensure a successful use case development. The three partners are responsible for the end-user perspective, infrastructure development, and interface implementation. The end-user of the system are researchers that perform fish track reconstruction science.

The end-user is represented by the [Institut Français de Recherche pour l'Exploitation de la Mer](#) (IFREMER). IFREMER is a consortium partner of the GFTS project. This institute is at the forefront of fish track research and has contributed to the definition of this use case from the proposal writing stage. They will use the GFTS as a platform for advancing their fish tracking modeling.

For infrastructure development and deployment the project has the [Simula Research Laboratory](#) (SRL) as a consortium partner. Simula is a research laboratory that performs high-quality information and communication technology research. It is the perfect partner for ensuring a highly scalable fish track reconstruction environment that IFREMER will rely on.

To create a decision support tool the project has [Development Seed](#) as a consortium partner. Development Seed is a pioneering technology company specializing in data visualization, mapping, and software development for social good, recognized for its innovative use of geospatial data to empower organizations and governments worldwide. Development Seed is also the lead of the project.

Use Case Description

Our use case has two main components, one is a modeling environment for fish track reconstruction, and the other is a decision support tool to explore future habitat conditions for fish species based on the reconstructed tracks.

Fish track modeling environment

The modeling environment addresses a need from the scientific community to have a consistent way of running fish track reconstruction at scale. Currently, scientists that have biologging data from in situ measurements have to perform fish track reconstruction on their own local machines. Fish track reconstruction is a highly resource intensive process. It requires access to large datasets of sea temperature profiles along time and space, as well as a high number of computing power to run the models. This is a limitation for biologists who are specialized in marine biology and biologging from in-situ fish tagging. These scientists may not have the knowledge of creating and maintaining scalable computing environments for fish track reconstruction. Furthermore, the ocean temperature data used for fish track reconstruction is shared across different modeling efforts for different species. A modeling environment that combines data availability with computational resources will significantly lower the barriers for biologists to perform fish track reconstruction. This is a clear value added when compared to the current workflow that biologists use for fish track reconstruction.

Our use case will therefore add value by creating a ready to use environment for fish track reconstruction from biologging data. This environment will be made available to scientists that are limited in their current workflow, reducing the barrier of entry for this kind of scientific analysis, and increasing the uptake of ocean temperature data created in the DestinE project and the DestinE Platform environment.

Preconditions

The first precondition for using the fish track modeling environment is familiarity with the pangeo-fish software and fish track reconstruction techniques. The model environment will be a ready-to-use development environment, but the user will still be expected to be an expert in fish track reconstruction. The environment is meant to relieve the burden of infrastructure management for users that know how to run models in notebooks.

A second precondition for using this system is the availability of biologging data. The system does not provide the biologging data itself. A user of the system will have to provide this data and upload it to the system prior to start modeling.

A third precondition is availability of computing resources. The DestinE Platform infrastructure will provide the underlying infrastructure. The requirements on the size of the computing resources will depend on the

amount of biologgin data that is available (how many species, how many observation points, for how many dates), as well as the resolution of the ocean temperature data used for reconstruction. Here the higher the better, each user will choose the best available data for the region where the biologing data was collected and the time range for which it is available.

Input summary

The inputs for this system are biologging data from fish species on the one hand, and ocean temperature on the other hand. The biologging data needs to be provided by the user, whereas the ocean temperature data will be available through the DestinE Platform data lake. This is one of the key advantages of this system, that the big data requirements for this kind of modeling are readily available through the DestinE Platform.

Output summary

The outputs of the fish track modeling environment are the reconstructed fish tracks. These are georeferenced fish locations with dates attached to them, as well as daily posterior distributions of fish position. These represent daily maps of fish presence probability. These fish tracks and location probabilities are stored within the GFTS system and can be integrated into the decision support system in a later step. The presence probability can be aggregated over time intervals such as spawning periods and individual fishes to help managers designing spatial management measures.

Workflow description

The high level workflow for using this system is to register on the platform. To control the type and level of usage of the platform, users will be able to join the GFTS platform by invitation only in the first iteration. After registration and login, the user will have access to a Jupyter environment that will scale on demand. The user will then upload their biologging data, perform modeling and store the output of the modeling in secure and long term cloud storage.

Future conditions decision support tool

Fish track reconstruction is a significant first step in better understanding fish populations in the oceans. However, the fish tracks themselves are not enough for decision makers to derive actionable insights for improving marine life protection and design fisheries policies.

The second part of the GFTS project is therefore to leverage the fish tracks and combine them with long term ocean temperature forecasts from the DestinE Climate Adaptation Digital Twins. We will calculate the future conditions of these fish track locations relying heavily on DestinE Digital Twin data and the core functionalities of the DestinE Platform. Ocean condition variables from the climate adaptation Digital Twin

forecasts will be crossed with reconstructed fish tracks using the scalable DestinE Platform. Example variables for these calculations are sea temperature and salinity.

This integration will enable us to evaluate a range of potential scenarios. These “what-if” scenarios will depict how variations in climate conditions could impact the quality of habitat for fish species based on the estimated fish track locations.

Preconditions

The precondition for providing the decision support tool for any fish species, is its previous completion of fish track reconstruction in the modeling environment. Once fish tracks are available, the system can integrate them and create what-if scenarios based on long term climate projections.

Another precondition is the existence of detailed data from the different Climate Adaptation Digital twin models, ideally under different climate scenarios. These are required to compute the what-if scenarios that analyze the future conditions that fish will meet in their current location.

Input summary

The input for the decision support system are the fish tracks, and future projections of ocean temperatures. With these two inputs, scenarios will be computed and made available to the user.

Output summary

The output of the decision support system are maps and graphs of future habitat conditions for the species at hand. The system will expose the what-if scenarios to the end user through an intuitive interface. Outputs are the graphical representation on the interface, as well as potentially options to export the underlying data.

Workflow description

The decision support system will be publicly available. The workflow begins therefore with accessing the tool through a browser. The user can then explore different scenarios for the available species in the interface. The detail of how this interaction will look like will be part of a co-creation process with our initial users and the development team.

Application to Seabass

To demonstrate the capabilities of the GFTS system, we will focus on Seabass, a species that has been studied by our scientific team for decades. We will leverage an extensive fish track biologging dataset that IFREMER has collected from 2010 onwards. The dataset is for Seabass along the French Atlantic coast.. The

combination of this dataset with the Climate Adaptation Digital Twin will allow us to develop a fully fledged usage of the GFTS infrastructure and demonstrate the functionality of the system. If data availability allows, we will also add other fish species such as Pollock to the analysis to demonstrate

Expected outcome

We expect the outcome of this project to support both the scientific community and policy makers in the space of marine biology and marine area protection. In a fast changing world it is important to ensure that scientists have the tools to scale their scientific methods to regional and global levels, so that the large-scale implications of scientific insights can be estimated. Once data and insights are generated, it is equally important to ensure that policy makers have access to this data in a way that is useful for them to maximize the positive outcome of the scientific output.

In the GFTS project we address the needs of the scientists and the policy makers at the same time. For this, we build a scalable fish track reconstruction environment for biologgging experts on the one hand, and help marine protection policy making by estimating impact of climate change on fish habitat conditions on the other hand. We hope this will contribute to an increased understanding of fish movement in our seas as well as some of the implications of climate change on fish populations based on the known fish movements.

The table below lists institutes that expressed interest in using the GFTS system for fish track analysis.

Institute	Species	Location
Swedish University of Agricultural Sciences	Shark	Sweden
Universidade do Algarve	-	Portugal
Dalhousie University	-	Canada
Flanders Marine Institute	-	Belgium
Nihon University College of Bioresource Sciences Department of Marine Science	-	Japan
Centre d'Etudes Biologiques de Chizé	-	France
Institute of Marine Research	-	Norway

Uncertainty

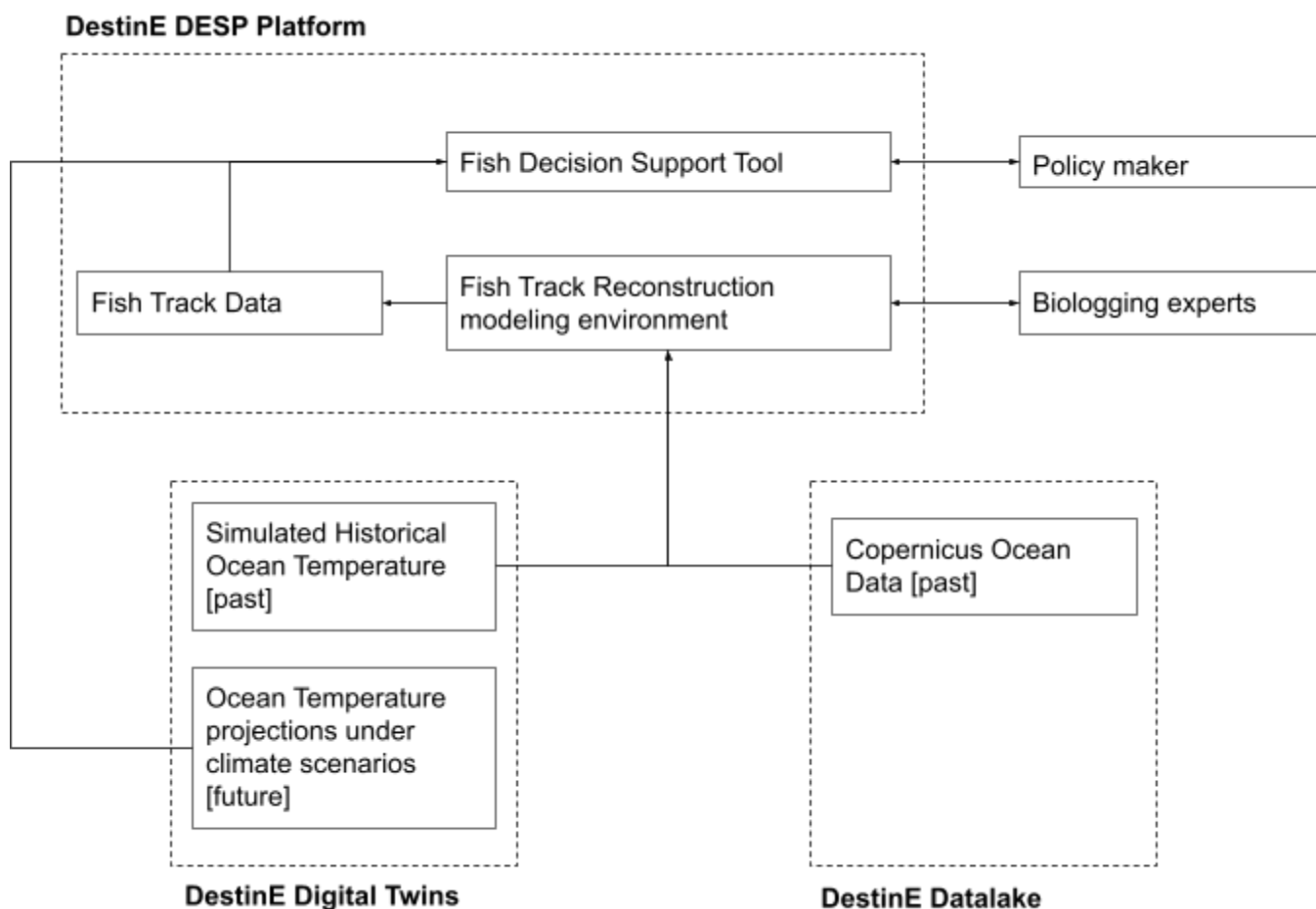
Our approach quantifies uncertainties using hidden Markov models to derive the posterior probability of the sequence of fish positions, from which by-products are derived as mean, maximum and most probable

fish trajectories. Appropriate ocean temperature and pressure datasets together with biologging in-situ datasets are used to estimate fish habitats, along with sensitivity analysis for a range of outcomes. The developed interactive tools offer end-users the ability to generate probabilistic predictions and explore different scenarios. Sensitivity analysis identifies key uncertainties and their impact on decisions, while expert judgment complements limited data. Subsequently, policy-making prioritizes robust strategies to design fish conservation areas.

Architecture

The technical framework of this proposal is based on the Pangeo ecosystem, which facilitates co-creation and solution development. An interactive and scalable Pangeo computing infrastructure will be deployed to provision the resources required for running the pangeo-fish model. We will connect all available and relevant ocean temperature and pressure datasets of DestinE using Pangeo techniques such as Intake, STAC and kerchunk. Then, in consultation with IFREMER's Ocean Physics scientist, additional datasets from IFREMER such as OSI-SAF datasets, and Copernicus marine services will be connected to the Pangeo DestinE Platform.

The following diagram shows an overview of what data is used as input, what data is being generated, and how this information feeds into the decision support tool.



Co-design

The co-creation process developed within the [e-shape project](#) will be adopted to ensure that all stakeholders are engaged and their needs are addressed. We will employ co-creation as a collaborative process where stakeholders and end-users actively engage in the creation of the GFTS platform. We are planning to reach out to potential users and help them to

1. Use the GFTS system and pangeo-fish software to run fish track reconstruction
2. Include their existing fish track data in the decision support tool
3. Explore early versions of the decision support tool and get feedback on how to advance

To achieve this, we will set up online workshops, interviews, and training sessions where relevant. The goal of these activities will be for the new users and stakeholders to become familiar with the GFTS system and be able to use it for their own purposes.

In general, we will encourage all participants to openly share ideas, and to joint decision-making to generate an innovative and tailored outcome for GFTS. By integrating diverse perspectives, knowledge, and

experiences, our co-creation process will ensure a solution that is more relevant, usable, and sustainable. This approach fosters a sense of ownership and commitment among stakeholders, leading to an outcome that better meets the needs and preferences of everyone involved.

DestinE Platform integration

This section contains the specification of how we will integrate with DestinE Platform at a high level. The integration plan details will evolve as we get onboarded to the DestinE Platform in practice. We will document the process of the integration of our use case into DestinE Platform in our public GitHub repository. Documenting our learnings as part of the first cohort of DestinE Platform users will hopefully help future use cases and other future DestinE Platform users to integrate their project with DestinE Platform easier and faster.

We will leverage the DestinE Platform wherever possible, from authentication, to existing Kubernetes clusters, object storage, and more. The use case will be made available to selected DestinE users and integrated in the DestinE Platform system native capabilities. The selected DestinE users can make use of the data available on the DestinE Platform, the DestinE data lake, and from the DestinE Digital Twins. Authorized users will be able to run the tools & services developed within the Use Case.

Interactive dashboards will be built based on Jupyter and Pangeo technologies such as Voila dashboards, Jupyterlite or Observable community notebooks. For the development phase, Pangeo JupyterHub with Dask cluster would be deployed on the DestinE Platform.

Planning and Strategy

Scalability Plan

We will build a scalable infrastructure from day one. The pangeo deployment will be based on Kubernetes and have an auto-scaling feature that will add more capacity to the infrastructure on demand as usage increases. Our service will rely on the ability of the DestinE Platform to provide the necessary resources to our Kubernetes cluster. The processing we use for the Seabass species will be a great scale test, as it is one of the larger existing biologging datasets in the world.

Long Term Strategy

Our long term strategy is to create a decision support tool that supports decision making around fish stocks around the globe. We believe that this requires the ability to estimate and map fish locations, primarily

focused on biologging data, but with the vision of extending to other sources that indicated potential fish habitat. This is why the GFTS system has two main components: the fish track reconstruction environment, and the decision support tool.

The fish track reconstruction itself is an expensive analysis and will require funding for data analysis on demand using a case by case basis. The fish track reconstruction environment is not freely accessible and we will only scale up for the resource intensive processes if there is funding for such additional analysis. When no processing is being performed, the environment will only require minimal resources.

The decision support tool will be kept online in the medium term. Also this system is expected to have low resource utilization. New data from additional analysis will only be integrated into the decision making tool when the fish track reconstruction mechanism is deployed for additional processing as described above.

Traceability

We are using ORCID IDs for all the work published under the GFTS project. We are also publishing the source code and the relevant documentation in the public [DestinE_ESA_GFTS GitHub repository](#). The documentation will be integrated into the codebase and published in a rendered version in a separate [url for documentation](#).

All the software developed within the project will be licensed through the Apache 2 OSI license and will be made available in Github. All developments will be done in the GitHub repository linked below and software releases will get a DOI in Zenodo to increase the Findability, Availability, Interoperability and Reusability (FAIR) of all the software components. FAIR [Research Objects](#) (ROs) will be created using the [RoHub](#), a Research Object Management platform in order to align with Open Science principles such as the practice of sharing inputs, outputs, models, software, workflows, brochures, training, and any other publication during the active phase of the project.

Link to public GitHub repository: https://github.com/destination-earth/DestinE_ESA_GFTS

Link to documentation website: https://destination-earth.github.io/DestinE_ESA_GFTS/

The use of commercial software is not foreseen. All the technologies are based on the Jupyter and Pangeo ecosystem that are widely adopted standard-based technologies enabling interoperability of DestinE components with external systems and guaranteeing compliance with standards indicated in DestinE Platform.

Use case associations

This section we will document any associations with other DestinE Platform use cases. There are no other use cases associated with the GFTS use case at the moment.

Use Case Notes

This section will contain any additional notes and information on the use case that users need to be aware of while using the suggested workflow. There are no additional use case notes at the moment.

GANTT Chart

Task number, focus area, Responsible leaders	WP1	WP2	WP3	WP4
Task 1: Daniel Wiesmann(DS)				
Task 2: Technology, Benjamin Ragan-Kelley (SRL)				
Task 2: Science, Tina Odaka (IFREMER);				
Task 3: Tina Odaka (Ifremer)				
Task 4: User experience, Mathieu Woillez (IFREMER);				
Task 4: Technology, Daniel Wiesmann (DS)				